

RESEARCH

Open Access



3D visual discomfort prediction using low complexity disparity algorithms

Jianguo Chen^{1,2*}, Jun Zhou¹, Jun Sun¹ and Alan C. Bovik²

Abstract

Algorithms that predict the degree of visual discomfort experienced when viewing stereoscopic 3D (S3D) images usually first execute some form of disparity calculation. Following that, features are extracted on these disparity maps to build discomfort prediction models. These features may include, for example, the maximum disparity, disparity range, disparity energy, and other measures of the disparity distribution. Hence, the accuracy of prediction largely depends on the accuracy of disparity calculation. Unfortunately, computing disparity maps is expensive and difficult and most leading assessment models are based on features drawn from the outputs of high complexity disparity calculation algorithms that deliver high quality disparity maps. There is no consensus on the type of stereo matching algorithm that should be used for this type of model. Towards filling this gap, we study the relative performances of discomfort prediction models that use disparity algorithms having different levels of complexity. We also propose a set of new discomfort predictive features with good performance even when using low complexity disparity algorithms.

Keywords: Visual discomfort, Low complexity disparity calculation algorithms, 3D NSS, Uncertainty map

Abbreviations: GGD, Generalized Gaussian distribution; LCC, Linear correlation coefficient; MOS, Mean opinion score; NSS, Natural scene statistics; QA, Quality assessment; SAD, Sum-of-absolute difference; SROCC, Spearman rank order correlation coefficient; SSIM, Structural similarity; SVR, Support vector regression; S3D, Stereoscopic 3D

1 Introduction

The human consumption of stereoscopic 3D (S3D) movies and images has dramatically increased in recent years. 3D content can better allow the user to understand the visual information being presented, thereby enhancing the viewing experience by providing a more immersive, stereoscopic visualization [1]. However, stereo images that have low-quality content or shooting errors can induce unwanted effects such as fatigue, asthenopia, eye strain, headache, and other phenomena conducive to a bad viewing experience [2]. A large number of studies have focused on finding features (e.g., disparity, spatial frequency, stimulus width, object size, motion [3], and crosstalk effects) that can be reliably extracted from 3D images (stereopairs) towards creating automatic 3D discomfort prediction algorithms to predict and potentially

reduce feelings of visual discomfort experienced when viewing 3D images [2, 4].

Several possible factors of visual discomfort have been extensively studied, such as the vergence-accommodation conflict [5, 6], excessive disparities and disparity gradients [7], prolonged viewing, the viewing distance [8], and the amount of defocus-blur [9]. Prolonged exposure to conflicts between vergence and accommodation is a main determinant of the degree of experienced visual discomfort and fatigue when viewing S3D content [9–11]. Hence, several predictive models have been built to simulate and predict occurrences of this phenomenon. Commonly, the features used in discomfort prediction models were extracted from disparity maps. These features included the disparity location, disparity gradient, disparity range, maximum angular disparity, and disparity distribution [7, 12–16]. Hence, the predictive powers of these discomfort assessment models strongly depends on the accuracy of disparity calculation.

However, there is no consensus regarding the type of disparity calculation algorithm that should be used for 3D visual discomfort. Early on, some developers used stereo

*Correspondence: skychenyuran@sjtu.edu.cn

¹Institute of Image Communication and Network Engineering, Shanghai Jiao Tong University, 200240 Shanghai, China

²Laboratory for Image and Video Engineering (LIVE), The University of Texas at Austin, 78701 Austin, USA

matching algorithms that extract only sparsely distributed disparities (e.g., at luminance edges) to achieve low complexity, fast computation [13, 14]. More, recent studies have emphasized the use of high complexity dense stereo matching algorithms that deliver high quality disparity maps, such as the matching algorithm [17] used in [7], dynamic programming [15, 18], the Depth Estimation Reference Software [19] used in [12], and combinations of sparse and dense disparity estimation methods [16].

Although high complexity dense disparity calculation algorithms deliver more accurate disparity results, speed of computation is desirable in many settings, e.g., on real-time 3D videos. However, there is scarce literature on the performance differences of 3D discomfort prediction models deploying different disparity algorithms nor of the causative factors contributing to these differences, such as complexity. Furthermore, little attention has been paid to balancing speed against prediction accuracy by making use of low complexity disparity algorithms. Towards filling these gaps, we begin by studying the performance differences of 3D discomfort prediction models using three nominal disparity algorithms having different levels of complexity. We then introduce two new sets of discomfort predictive features, the uncertainty map and natural scene statistics, which have previously found use in 3D image quality assessment models [20–22]. These features efficiently improve the performance of prediction models that use low complexity disparity calculation methods.

2 Background

The main difference between viewing natural scenes and viewing a stereoscopic display is that vergence and accommodation normally occur in a synergistic manner in natural viewing but they do not when viewing a display. In a 3D scene viewed on a stereoscopic display, accommodation is fixed by the distance of the dichoptic images from the two eyes but vergence is free to adapt to the disparity-defined depth planes that occur when a fused image is achieved. This perceptual conflict is a main cause of visual discomfort. As the binocular disparity signal is the primary cue in evoking vergence [23], extracting accuracy disparity signals from stereoscopic image pairs is the first important step to make good predictions of the degree of visual discomfort experienced when viewing 3D images.

Stereo matching is the most common method to extract disparity signals from image pairs. The disparity signals (in pixels) which are extracted by stereo matching algorithms can be converted to retinal disparities (in angles) given the viewing parameters and the size of the display [24]. Although this conversion is not linear, most studies prefer to using pixel disparities when conducting visual discomfort modeling to simplify algorithm design [7, 13–16, 25]. We will also use pixel disparity-based features.

Research on stereo matching algorithm design has been a topic of intense inquiry for decades. Stereo matching algorithms can be classified into sparse and dense stereo matching. Sparse stereo matching methods do not calculate disparity at every pixel and are deployed for their low complexity or if only sparse data is needed. Dense stereo matching methods calculate disparity at every pixel. Most recent discomfort assessment models are built on dense stereo matching algorithms [26].

All dense stereo matching algorithms use some method of measuring the similarity of pixels between the two image views. Typically, a matching function is computed at each pixel for all disparities under consideration. The simplest matching functions assume that there is little or no luminance difference between corresponding left/right pixels, but more robust methods may allow for (explicitly or implicitly) radiometric changes and/or noise. Common pixel-based matching functions include absolute differences, squared differences, or sampling-insensitive absolute differences [27]. Common window-based matching functions include the sum of absolute or squared differences (SAD, SSD), normalized cross-correlation (NCC), and rank and census transforms [28]. Some matching functions can be implemented efficiently using unweighted and weighted median filters [29, 30]. More complicated similarity measures are possible and have included mutual information or approximate segment-wise mutual information as used in the layered stereo approach of Zitnick [31]. Some methods not only try to employ new combined matching functions but also propose secondary disparity refinement to further remove the remaining outliers [32].

In order to gain insights into the influence of the choice of stereo algorithm on the performance of 3D visual discomfort models, we selected three popular and characteristic dense stereo algorithms, ranging from a computationally expensive, high performance model (e.g., as assessed on the Middlebury database [33]) to a very simple, inexpensive model that delivers reasonable performance.

Researchers have deployed a wide variety of stereo matching algorithms to obtain disparity maps for assessment 3D discomfort prediction models [16–19]. The algorithms previously used are characterized by high computational complexity and generally deliver highly accurate disparity maps. Of the three disparity engines we use, the optical flow software (DFLOW) [17] delivers highly competitive predictions of disparity on the Middlebury Stereo Evaluation dataset [33]. This tool has been utilized in a mature 3D visual discomfort assessment framework which achieves good predictive power [7].

The second comparison algorithm is a window-based stereo matching algorithm based on the SSIM [34] index (DSSIM) [20]. The disparity map of a stereo pair is

generated by using SSIM as the matching objective, resolving ties by a minimum disparity criterion. This algorithm was used in a popular 3D QA model [20] but has not yet been utilized in previous 3D visual discomfort assessment models.

The third algorithm (DSAD) was chosen for its very low complexity. It uses a window-based sum-of-absolute difference (SAD) luminance matching functional without a smoothness constraint. This is a very basic stereo matching algorithm that has only been used in early, simple 3D visual discomfort prediction models.

3 Affect comparison of disparity estimation on visual discomfort prediction

Figure 1 shows four images (“cup,” “human,” “lawn,” and “stone”) from the IEEE-SA stereo image database [35] and disparity maps extracted by these three algorithms. Figure 2 are corresponding depth distribution histograms computed from the depth maps delivered by these three algorithms. The search range of DSSIM and DSAD was fixed at $[-120, 90]$ which is the maximum and minimum disparities of images in the IEEE-SA database. The values of the disparity maps range from dark to white denoting disparity ranging from maximum to minimum.

It is apparent that the disparity maps extracted by DFLOW yield the highest quality of depth detail. The disparity maps delivered by DSSIM are of much lower reliability than those of DFLOW. The disparity maps from DSAD are even worse than those of DSSIM. There are many areas with false disparities. Among the three methods DFLOW, DSSIM, and DSAD, there is a decreasing degree of coherence and segmentability of the computed disparity patterns. Often, disparity errors occur on complex textured regions which the lower complexity stereo algorithms handle less well.

Clearly, the DSSIM and DSAD disparity maps would be difficult to apply in 3D visual discomfort prediction frameworks that require depth segmentation. Hence, we instead only study discomfort prediction frameworks based on analysis of the disparity distribution. Four features are extracted based on the study in [7]. The first two features are the mean values of the positive and negative disparities. These are computed separately since it is known that the sign of disparity can affect experienced visual discomfort [13, 36]:

$$f_1 = \frac{1}{N_{\text{Pos}}} \sum_{D(n)>0} D(n) \quad (1)$$

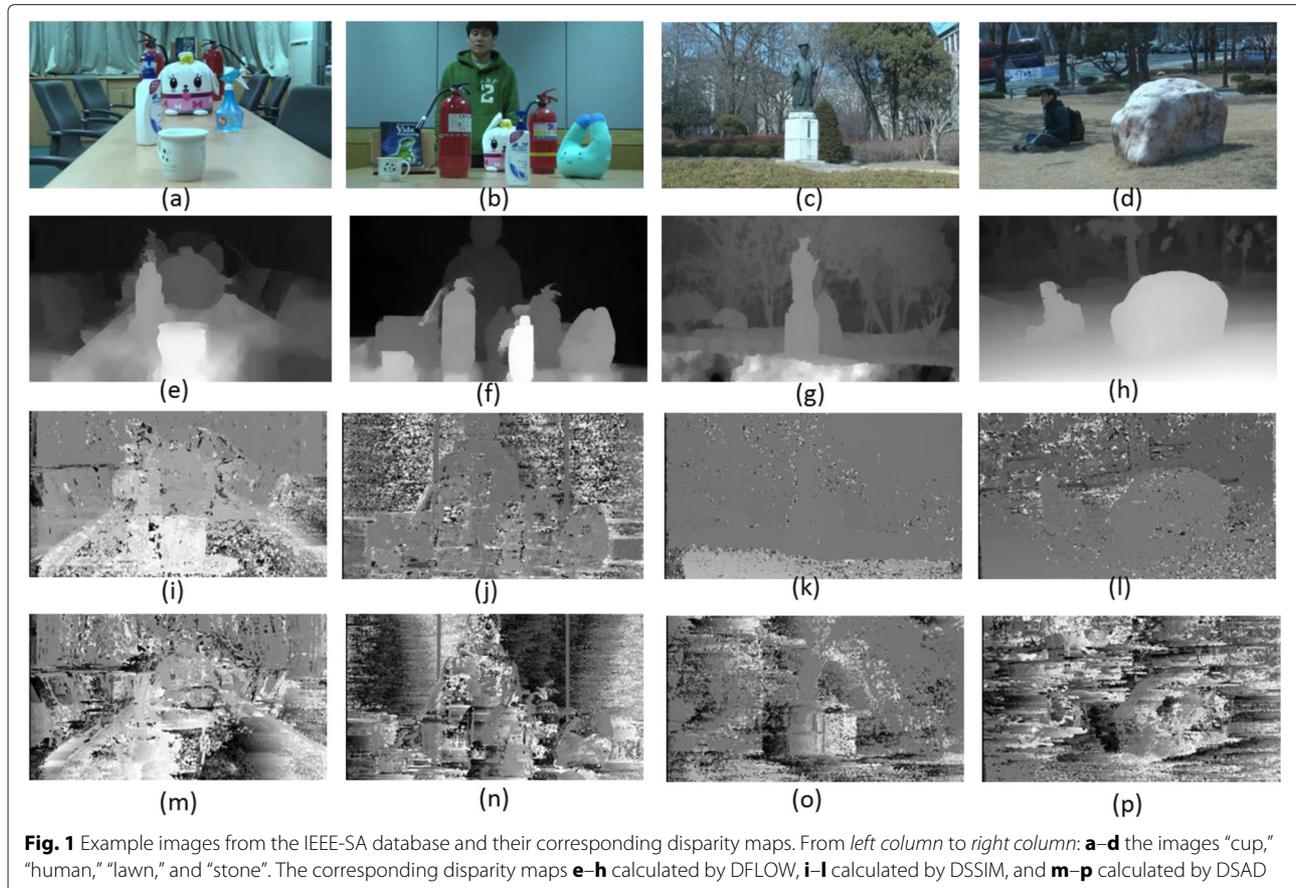
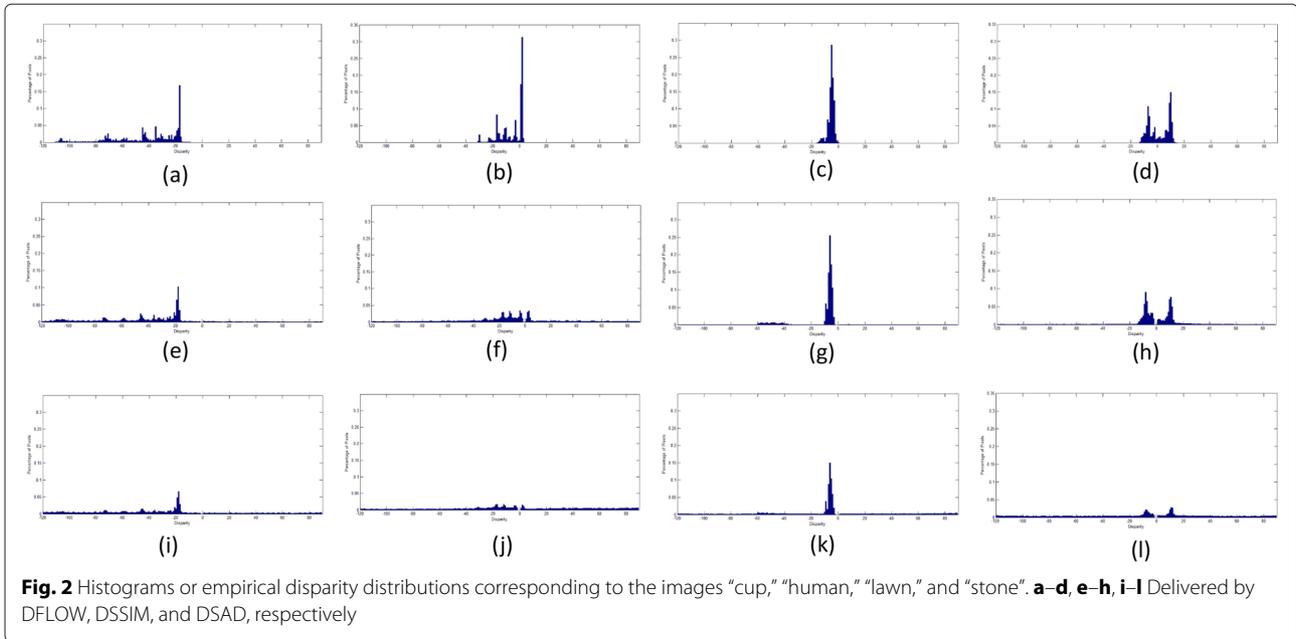


Fig. 1 Example images from the IEEE-SA database and their corresponding disparity maps. From left column to right column: **a–d** the images “cup,” “human,” “lawn,” and “stone”. The corresponding disparity maps **e–h** calculated by DFLOW, **i–l** calculated by DSSIM, and **m–p** calculated by DSAD



$$f_2 = \frac{1}{N_{\text{Neg}}} \sum_{D(n) \leq 0} D(n) \quad (2)$$

In (1) and (2), $D(n)$ is the n th smallest value in the disparity map, while N_{Pos} and N_{Neg} are the number of positive and negative values in the disparity map, respectively. If $N_{\text{Pos}} = 0$ or $N_{\text{Neg}} = 0$, then $f_1 = 0$ or $f_2 = 0$.

The average of the upper and lower 5 % disparities define the third and fourth features:

$$f_3 = \frac{1}{N_{5\%}} \sum_{n \leq N_{\text{total}} \times 5\%} D(n) \quad (3)$$

$$f_4 = \frac{1}{N_{95\%}} \sum_{n \geq N_{\text{total}} \times 95\%} D(n) \quad (4)$$

where $N_{5\%}$ and $N_{95\%}$ are the number of values that are lower and higher than 5 % and 95 % of the disparity values, respectively.

We extracted these four basic statistical features from disparity maps calculated by the three abovementioned stereo depth-finding algorithms on the stereo pairs in the IEEE-SA stereo image database [35]. The IEEE-SA stereo image database contains 800 stereo image pairs of high-definition (HD) resolution (1920×1080 pixels). An integrated twin-lens PANASONIC AG-3DA1 3D camcorder was used to capture the 3D content in the database. The subjective discomfort assessment experiment was conducted in a laboratory environment commensurate with standardized recommendations for subjective evaluation of picture quality [37]. A 46-in. polarized stereoscopic monitor of HD resolution was used to display the test stereo images. Each subject viewed the test stereo images

from a distance of about 170 cm, or about three times the height of the monitor. Twenty-four valid subjects participated in the subjective test. Each subject was asked to assign a visual discomfort score to each stereo test image on a Likert-like scale: 5 = very comfortable, 4 = comfortable, 3 = mildly comfortable, 2 = uncomfortable, and 1 = extremely uncomfortable. More information can be found in [25].

Simply stated, the images and corresponding MOS of these images were divided into test and training subsets. A support vector regression (SVR) was deployed as a regression tool on the training set and then applied to the test set. To implement the SVR, we used the LibSVM package [38] with the radial basis function kernel, whose parameters were estimated by cross-validation during the training session. One thousand iterations of the train-test process were applied where the image database was randomly divided into 80 % training and 20 % test at each iteration. The training and testing subsets did not overlap in content. The performance was measured using Spearman’s Rank Ordered Correlation Coefficient (SROCC) and (Pearson’s) linear correlation coefficient (LCC) between the predicted scores and the MOS. Higher SROCC and LCC values indicate good correlation (monotonicity and accuracy) with human quality judgments. We obtained the mean, median, and standard deviations of LCC and SROCC of the three models against MOS over all 1000 train-test trials, as tabulated in Table 1. Values of LCC and SROCC close to 1 mean superior linear and rank correlation with MOS, respectively. Obviously, the higher the mean and median, the better the LCC and SROCC performance. Conversely,

Table 1 Mean SROCC and LCC over 1000 trials of randomly chosen train and test sets on the IEEE-SA database

	SROCC _{Mean}	SROCC _{Med}	SROCC _{STD}	LCC _{Mean}	LCC _{Med}	LCC _{STD}
DFLOW	0.7445	0.7457	0.0389	0.8318	0.8358	0.0317
DSSIM	0.6628	0.6627	0.0426	0.7006	0.7019	0.0423
DSAD	0.5873	0.5889	0.0493	0.6057	0.6083	0.0491

a higher standard deviation implies more unstable performance.

From the results, we can see that the predictive power of the four-feature discomfort prediction models is dramatically reduced by the use of a low complexity stereo algorithm instead of a high performing, high complexity algorithm.

There is a significant increase in pixels having large estimated disparity errors in the disparity maps extracted by DSSIM and DSAD. By observing the histograms of the disparity distributions in Fig. 2, it may be seen that the disparities produced by DSSIM and DSAD span nearly the entire disparity range. Hence, it is difficult to obtain accurate values of the mean negative and positive disparities, nor the top 5 % biggest and smallest disparities. For example the four feature values (1)–(4) extracted by DFLOW on the image “human” were [1.69, -12.5, -26.9, 2.5], the values computed using DSSIM were [32.6, -33.5, -107.4, 78.8], and those using DSAD were [45.8, -45.0, -111.4, 85.5]. The largest and smallest 5 % of disparities found by DSAD essentially bracket the entire disparity.

Table 2 compares the computation times and estimation accuracies of these three disparity calculation methods. The computation times were recorded in units of hours on the IEEE-SA database. Since IEEE-SA does not provide ground truth maps, the estimation accuracies of these three algorithms were tested on the Middlebury stereo database [33]. The average percentages of bad pixels was recorded for each algorithm. From Table 2, it is apparent that the DSAD disparity algorithm executes with the fastest computation speed but it achieves the worst estimation accuracy.

Feature extraction from disparity distributions measured on the DSSIM and DSAD maps will likely be seriously affected by the high percentages of estimated errors, thereby adversely affecting discomfort prediction results. This would seem to advocate the use of only high complexity, high performance stereo modules in 3D visual

discomfort prediction models. However, another possibility worth exploring to improve the usability of disparity maps extracted by low complexity algorithms like DSAD or DSSIM, is to develop additional resilient features on them that can ameliorate the effects of disparity estimation errors.

4 Uncertainty map

A promising approach is to understand the distribution of estimated errors, from which useful features may be developed to improve the performance of discomfort prediction models using low-complexity stereo algorithms.

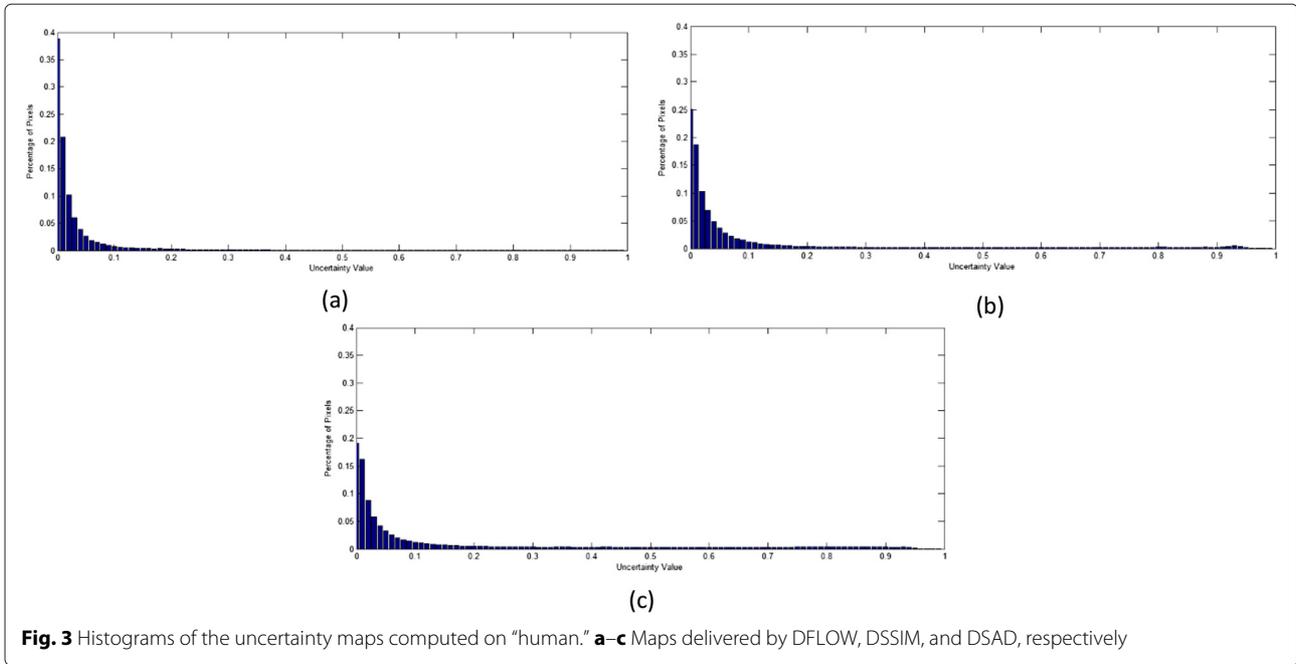
Pixels associated with disparity errors are often dissimilar with features computed on the corresponding disparity shifted pixels in the other view. The authors of [39] defined a disparity uncertainty map to estimate the uncertainty produced by DSSIM and used it as a feature to improve the task of 3D no-reference distortion assessment. The uncertainty is defined as:

$$\text{Uncertainty}(l, r) = 1 - \frac{(2\mu_l\mu_r + C_1)(2\sigma_{lr} + C_2)}{(\mu_l^2 + \mu_r^2 + C_1)(\sigma_l^2 + \sigma_r^2 + C_2)} \quad (5)$$

where l is the left-view image and r is the disparity-compensated right-view image of a stereo pair, μ and σ are the local weighted mean and weighted standard deviation computed over a local Gaussian window, and $C = 0.01$ is a constant that ensures stability. An 11×11 Gaussian weighting matrix with a space constant 3.67 pixels was used to compute μ and σ as in [39]. The uncertainty reflects the degree of similarity between the corresponding pixels of a stereo pair. Hence, the uncertainty distribution of a disparity map can be used to represent the distribution of estimated errors. Figure 3 shows the uncertainty distributions of DFLOW, DSSIM, and DSAD maps computed on the image “human.” It may be observed that the histogram computed on the DFLOW uncertainty map corresponds to a very peaked distribution. The histograms of the DSSIM and DSAD uncertainty maps are less peaky since more large estimated errors occur. This is consistently the case for the distributions of DFLOW, DSSIM, and DSAD maps on the other images in the IEEE-SA database. This phenomenon may be understood by observing that the stereo matching algorithms find good matches (with low uncertainty) at most places, while less common occluded or ambiguous flat or textured areas may cause sparse disparity errors (with high uncertainty).

Table 2 Compute times and accuracies of disparity calculation algorithms

	DFLOW	DSSIM	DSAD
Time/hour	45.71	22.04	3.51
Average percent of bad pixels	15.87 %	29.47 %	66.03 %



A log-normal distribution can be fit to the histogram of the uncertainty map [39]. The probability density function of a log-normal distribution is:

$$l_x(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp - \frac{(\ln x - \mu)^2}{2\sigma^2} \quad (6)$$

where μ is the location parameter and σ is the scale parameter. A simple maximum likelihood method can be used to estimate μ and σ for a given histogram of uncertainties [39].

To summarize, the features used to describe estimated disparity errors are the best-fit log-normal parameters (μ and σ), and the sample skewness and kurtosis of the uncertainty map which are calculated as (7) and (8):

$$s = \frac{\sum_{(i,j)} (U_{(i,j)} - \bar{U})^3 / N}{\sigma_U^3} \quad (7)$$

$$k = \frac{\sum_{(i,j)} (U_{(i,j)} - \bar{U})^4 / N}{\sigma_U^4} \quad (8)$$

where $U_{(i,j)}$ is the uncertainty value at coordinate (i, j) , \bar{U} is the mean, σ_U is the standard deviation, and N is the number of pixels.

5 3D NSS model

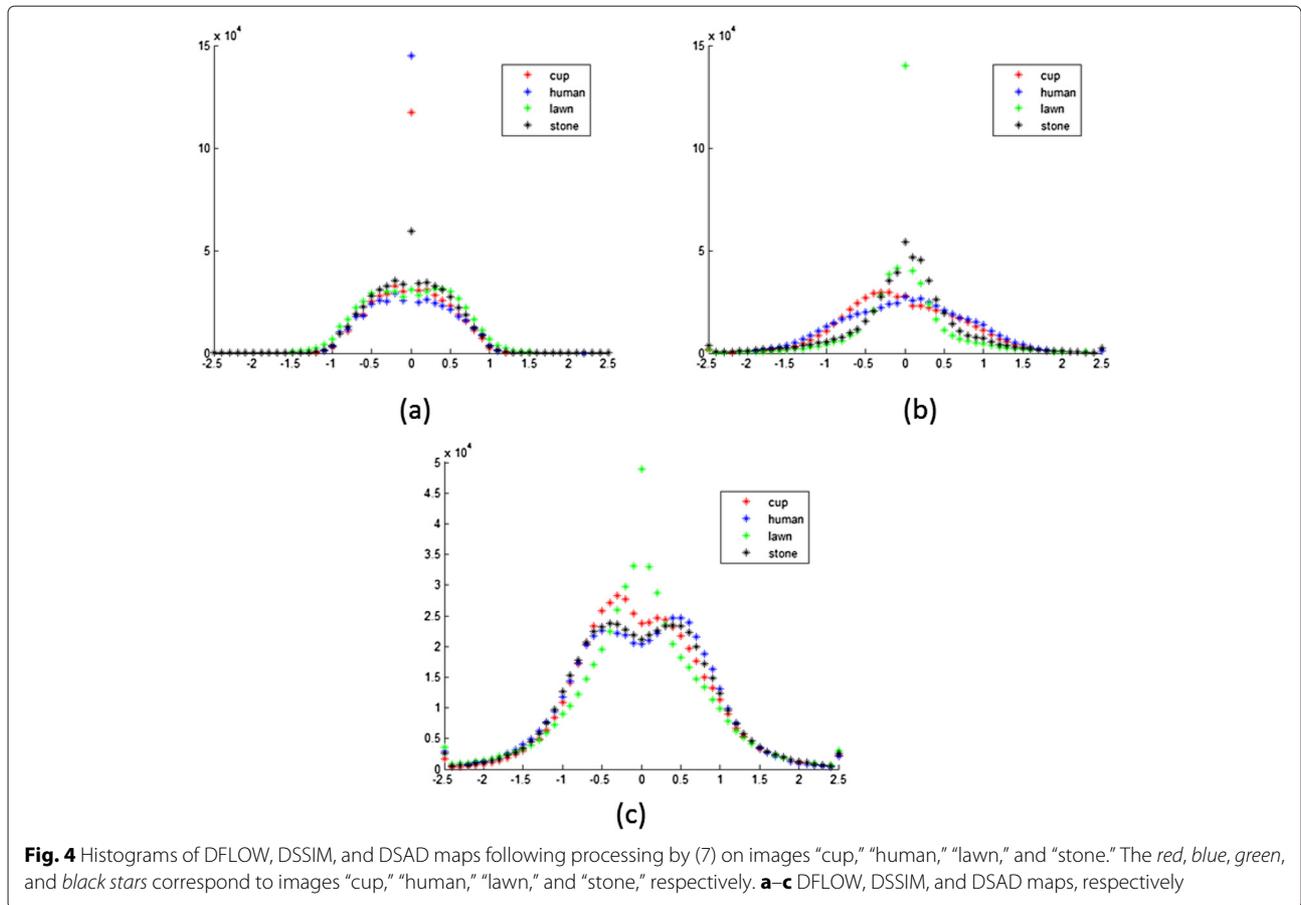
Towards ameliorating the weaknesses introduced by the use of low-complexity stereo models, we take a statistical approach towards characterizing the errors introduced by these algorithms. We accomplish this by subjecting the

computed disparity maps to a perceptual transform characterized by a bandpass process followed by a nonlinearity. The resulting data are then amenable to analysis under a simple but powerful natural scene model. Research on natural scene statistics (NSS) has clearly demonstrated that images of natural scenes belong to a small set of the space of all possible signals and that they obey predictable statistical laws [40]. Further, the studies of Hibbard [41] and Liu [42] found that the distribution of disparity follows a Laplacian shape. The authors of [39] processed the depth and disparity maps by local mean removal and divisive normalization and found that the histograms of the processed depth and disparity maps take a zero-mean symmetric Gaussian-like shape. One form of this process is [43]:

$$M(i, j) = \frac{I(i, j) - \mu(i, j)}{\sigma(i, j) + C} \quad (9)$$

where i, j are spatial indices, μ and σ are the local weighted mean and weighted standard deviation computed over a local Gaussian window, and $C = 0.01$ is a constant that ensures stability. An 11×11 Gaussian weighting matrix with a space constant 3.67 pixels is used to compute μ and σ as [39].

We applied the identical process (9) to DSAD, DSSIM, and DFLOW maps. The processed histograms for each computed on image "cup," "human," "lawn," and "stone" are shown in Fig. 4a-c. All of the histograms computed from DFLOW maps take zero-mean symmetric Gaussian-like shape as elaborated in [39]. Most of the histograms computed on DSSIM maps also take the same shape, but the modes of a few of them are shifted (e.g., "cup"). Other



than image “lawn,” the histograms of images processed by DSAD then subjected to DSAD disparity extraction fail to take a symmetric Gaussian-like shape. As in [39], when the Gaussian model fails, a generalized Gaussian distribution (GGD) fit may be attempted:

$$g_x(x; \mu, \sigma^2, \gamma) = ae^{-[b|x-\mu|]^\gamma} \quad (10)$$

where μ , σ^2 , and γ are the mean, variance, and shape-parameter of the distribution,

$$a = \frac{b\gamma}{2\Gamma(1/\gamma)} \quad (11)$$

$$b = \frac{1}{\sigma} \sqrt{\frac{\Gamma(3/\gamma)}{\Gamma(1/\gamma)}} \quad (12)$$

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

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt, x > 0 \quad (13)$$

The parameters (σ and γ) are estimated here using the method used in [44].

The authors of [39] use the GGD parameters (μ and σ), along with the sample standard deviation, skewness, and kurtosis of these coefficients as 3D features to estimate the quality of 3D images. Here, we deploy the same features to model a perceptually processed disparity distribution. Since the histograms of perceptually processed low quality disparity maps extracted by low complexity stereogram algorithms such as DSSIM or DSAD do not result in very

Table 3 Mean SROCC and LCC over 1000 trials on DSAD-based discomfort predictor

	SROCC _{Mean}	SROCC _{Med}	SROCC _{STD}	LCC _{Mean}	LCC _{Med}	LCC _{STD}
UM	0.6793	0.6813	0.0422	0.7067	0.7102	0.0417
NSS	0.6678	0.6699	0.0423	0.6959	0.6964	0.0402
UM+NSS	0.6492	0.6502	0.0519	0.7277	0.7286	0.0456

Table 4 Mean SROCC and LCC over 1000 trials on DSSIM-based discomfort predictor

	SROCC _{Mean}	SROCC _{Med}	SROCC _{STD}	LCC _{Mean}	LCC _{Med}	LCC _{STD}
UM	0.7102	0.7126	0.0367	0.7424	0.7422	0.0355
NSS	0.6981	0.6983	0.0396	0.7575	0.7608	0.0349
UM+NSS	0.7307	0.7306	0.0360	0.7853	0.7847	0.0341

good fitting results, then the average GGD fitting error is extracted as a useful feature:

$$\varepsilon = \frac{1}{N} \sum_x |H(x) - f_x(x)| \quad (14)$$

where N is the number of distributions in histogram, $H(x)$ is the quantity of pixels at coordinate x , and $f_x(x)$ is the fit result of GGD.

6 Performance evaluation

To summarize our model, we have devised two kinds of features that are designed to improve the prediction performance of 3D visual discomfort model that rely low complexity disparity calculation algorithms. These features are the uncertainty map (UM) which simulates estimated disparity errors; the best-fit log-normal parameters (μ and σ), skewness, and kurtosis of the uncertainty map; and 3D NSS features that serve as a prior constant on true disparity including the GGD parameters (μ and σ), standard deviation, skewness and kurtosis of perceptually processed disparity maps, along with the average GGD fitting error.

The testing that was done is similar to what was described earlier, but using combinations of these new features. The test was conducted on the IEEE-SA stereo image database [35], SVR was deployed as the regression tool, 1000 iterations of the train-test process were used, and image database was randomly divided into 80 % training and 20 % test sets. The performance was measured using SROCC and LCC between the predicted scores and the MOS. The operation environment was an Apple computer running Matlab: MacPro 4.1 with Intel xeon cpu e5520 2.27 Ghz and 6 GB of RAM.

Several combinations of the features are selected: UM, NSS, and (UM+NSS) integrated into the existing prediction framework. Three same disparity calculation were used.

We obtained the mean, median, and standard deviations of LCC and SROCC of the performance results of these combinations of features against MOS over all 1000 train-test trials, as tabulated in Tables 3, 4, and 5 for DSAD, DSSIM, and DFLOW, respectively. Table 6 shows the performance results of these combinations without considering the features from disparity. The performance results of prior models are also tabulated in Table 5. We tested the models contributed by Park [7], Nojiri [13], Yano [14], Choi [15], and Kim [16].

From Table 6, it may be observed that 3D NSS and the UM are predictive of the degree of visual discomfort induced by 3D images.

By observation of Tables 1 and 3, both kinds of features contribute to improving the performance of the nominal discomfort prediction framework using DSAD. In terms of mean SROCC, it is increased significantly from 0.5873 to 0.6793 using UM, and to 0.6678 using NSS. The combination of these features achieves the best results with mean SROCC of 0.7100 and LCC of 0.7314. These results are better than those of Nojiri [13], Yano [14], Choi [15], and Kim [16], and close to Park [7]. The stability of the predictive power is also improved in regard to the standard deviation of SROCC, 0.0493 to 0.0366.

A similar result is attained when using the DSSIM algorithm. The combination of features improves the performance of the prediction framework from SROCC 0.6628 to 0.7307 which is better than the result attained on DSAD. The stability is improved too.

Table 5 Mean SROCC and LCC over 1000 trials on DFLOW-based discomfort predictor and prior methods

	SROCC _{Mean}	SROCC _{Med}	SROCC _{STD}	LCC _{Mean}	LCC _{Med}	LCC _{STD}
Nojiri [13]	0.6108	0.6155	0.0732	0.6854	0.6935	0.0788
Yano [14]	0.3363	0.3384	0.0732	0.3988	0.4045	0.0748
Choi [15]	0.5851	0.5909	0.0798	0.6509	0.6565	0.0703
Kim [16]	0.6151	0.6195	0.0700	0.7018	0.7113	0.0771
Park [7]	0.7831	0.7882	0.0451	0.8604	0.8672	0.0482
UM	0.7626	0.7646	0.0355	0.8408	0.8437	0.0332
NSS	0.7862	0.7883	0.0322	0.8585	0.8594	0.0255
UM+NSS	0.8011	0.8064	0.0354	0.8649	0.8667	0.0285

Table 6 Mean SROCC and LCC over 1000 trials without considering the features from disparity

	SROCC _{Mean}	SROCC _{Med}	SROCC _{STD}	LCC _{Mean}	LCC _{Med}	LCC _{STD}
UM	0.4753	0.4801	0.0595	0.5418	0.5394	0.0615
NSS	0.6153	0.6219	0.0480	0.6806	0.6839	0.0388
UM+NSS	0.7100	0.7098	0.0366	0.7314	0.7323	0.0408

The new features also improve the performance of prediction framework based on the high complexity algorithm DFLOW, as shown in Table 7. Unlike the results on DSAD and DSSIM, here NSS contributes the most to the performance improvement. That may follow because the uncertainty map may not be able to improve the models much if the disparities are already accurately estimated. The contribution of NSS is stable over the visual discomfort models.

Table 7 shows the results of F-tests conducted to assess the statistical significance of the errors between the MOS scores and the model predictions on the IEEE-SA database. $(UM+NSS)_{DF}$ means the model with features of UM, NSS, and disparity using the DFLOW disparity calculation method. The residual error between the predicted score of a model and the corresponding MOS value in the IEEE-SA database can be used to test the statistical efficacy of the model against other models. The residual errors between the model predictions and the MOS values are:

$$R = \{Q_i - MOS_i, i = 1, 2, \dots, N_T\} \quad (15)$$

where Q_i is the i th objective visual discomfort score and MOS_i is the corresponding i th MOS score. The F-test was used to compare one objective model against another objective model at the 99.9 % significance level. Table 7 is the result of the F-test. A symbol value of “1” indicates that the statistical performance of the model in the row is better than that of the model in the column, while “0” indicates the performance in the row is worse than that in the column, and “–” indicates equivalent performance. The results indicate that both UM and NSS features improve the performances of the models with statistical significance.

Table 7 Results of the F-test performed on the residuals between objective visual discomfort predictions and MOS values at a significance level of 99.9 %

	Kim	Park	DFLOW	UM _{DF}	NSS _{DF}	$(UM+NSS)_{DF}$
Kim	–	0	0	0	0	0
Park	1	–	1	1	–	0
DFLOW	1	0	–	0	0	0
UM _{DF}	1	0	1	–	0	0
NSS _{DF}	1	–	1	1	–	0
$(UM+NSS)_{DF}$	1	1	1	1	1	–

Compared to the computation time of DSAD (3.51 h), the average computation time of these two features on the IEEE-SA database was much reduced (0.78 h). Hence, UM and NSS can efficiently improve visual discomfort models without much extra computation.

7 Conclusions

We studied the performance differences of 3D discomfort prediction models that rely on three disparity calculation algorithms having different complexity levels. The experimental results showed that the predictive power of a nominal prediction model is dramatically reduced when using a low complexity algorithm instead of a high complexity algorithm. The performance of models under the low complexity algorithm is also more unstable. Two kinds of new features were introduced to stabilize low-complexity results: features of a disparity uncertainty map (UM) and features of a 3D NSS model. We find that integrating these features significantly elevates the performance of the nominal discomfort model using low complexity stereo algorithms like DSAD or DSSIM. The new features also improve performance when a high complexity disparity estimator is used.

Acknowledgements

The work for this paper was supported by NSFC under 61471234, MOST under Contact 2015BAK05B03, 2013BAH54F04.

Authors' contributions

JC carried out this visual discomfort prediction model and drafted the manuscript. JZ and JS participated in the design of the study and performed the statistical analysis. AB conceived of the study, participated in its design and coordination, and helped to draft the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 23 October 2015 Accepted: 9 August 2016

Published online: 31 August 2016

References

- GR Jones, D Lee, NS Holliman, D Ezra, in *Photonics West 2001-Electronic Imaging*. Controlling perceived depth in stereoscopic images (International Society for Optics and Photonics, San Jose, 2001), pp. 42–53
- M Lambooj, M Fortuin, I Heynderickx, W IJsselsteijn, Visual discomfort and visual fatigue of stereoscopic displays: a review. *J. Imaging Sci. Technol.* **53**(3), 30201–1 (2009)
- J Li, M Barkowsky, P Le Callet, Visual discomfort of stereoscopic 3d videos: Influence of 3d motion. *Displays*. **35**(1), 49–57 (2014)
- M Lambooj, WA IJsselsteijn, I Heynderickx, Visual discomfort of 3d tv: Assessment methods and modeling. *Displays*. **32**(4), 209–218 (2011)

5. M Urvov, M Barkowsky, P Le Callet, How visual fatigue and discomfort impact 3d-tv quality of experience: a comprehensive review of technological, psychophysical, and psychological factors. *Ann. Telecommun-Ann. Télécommun.* **68**(11–12), 641–655 (2013)
6. R Patterson, Review paper: Human factors of stereo displays: An update. *J. Soc. Inf. Disp.* **17**(12), 987–996 (2009)
7. J Park, S Lee, AC Bovik, 3d visual discomfort prediction: vergence, foveation, and the physiological optics of accommodation. *IEEE J. Sel. Top. Sign. Process.* **8**(3), 415–427 (2014)
8. R Patterson, Human factors of 3-d displays. *J. Soc. Inf. Disp.* **15**(11), 861–871 (2007)
9. FL Kooi, A Toet, Visual comfort of binocular and 3d displays. *Displays.* **25**(2), 99–108 (2004)
10. M Emoto, T Niida, F Okano, Repeated vergence adaptation causes the decline of visual functions in watching stereoscopic television. *Disp. Technol. J.* **1**(2), 328–340 (2005)
11. T Shibata, J Kim, DM Hoffman, MS Banks, The zone of comfort: Predicting visual discomfort with stereo displays. *J. Vis.* **11**(8), 11–11 (2011)
12. H Sohn, YJ Jung, S-i Lee, YM Ro, Predicting visual discomfort using object size and disparity information in stereoscopic images. *IEEE Trans. Broadcast.* **59**(1), 28–37 (2013)
13. Y Nojiri, H Yamanoue, A Hanazato, F Okano, in *Electronic Imaging 2003*. Measurement of parallax distribution and its application to the analysis of visual comfort for stereoscopic hdtv (International Society for Optics and Photonics, Santa Clara, 2003), pp. 195–205
14. S Yano, S Ide, T Mitsuhashi, H Thwaites, A study of visual fatigue and visual comfort for 3d hdtv/hdtv images. *Displays.* **23**(4), 191–201 (2002)
15. J Choi, D Kim, S Choi, K Sohn, Visual fatigue modeling and analysis for stereoscopic video. *Opt. Eng.* **51**(1), 017206–1 (2012)
16. D Kim, K Sohn, Visual fatigue prediction for stereoscopic image. *IEEE Trans. Circ. Syst. Vi. Technol.* **21**(2), 231–236 (2011)
17. D Sun, S Roth, MJ Black, in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference On*. Secrets of optical flow estimation and their principles (IEEE, San Francisco, 2010), pp. 2432–2439
18. D Scharstein, R Szeliski, A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. J. Comput. Vis.* **47**(1–3), 7–42 (2002)
19. M Tanimoto, T Fujii, K Suzuki, N Fukushima, Y Mori, Depth estimation reference software (ders) 5.0. ISO/IEC JTC1/SC29/WG11 M. **16923**, 2009 (2009)
20. M-J Chen, C-C Su, D-K Kwon, LK Cormack, AC Bovik, Full-reference quality assessment of stereopairs accounting for rivalry. *Signal Process. Image Commun.* **28**(9), 1143–1155 (2013)
21. AC Bovik, Automatic prediction of perceptual image and video quality. *Proc. IEEE.* **101**(9), 2008–2024 (2013)
22. M-J Chen, LK Cormack, AC Bovik, No-reference quality assessment of natural stereopairs. *IEEE Trans. Image Process.* **22**(9), 3379–3391 (2013)
23. AM Horwood, PM Riddell, The use of cues to convergence and accommodation in naïve, uninstructed participants. *Vis. Res.* **48**(15), 1613–1624 (2008)
24. T-H Lin, S-J Hu, Perceived depth analysis for view navigation of stereoscopic three-dimensional models. *J. Electron. Imaging.* **23**(4), 043014–043014 (2014)
25. J Park, H Oh, S Lee, AC Bovik, 3d visual discomfort predictor: Analysis of disparity and neural activity statistics. *IEEE Trans. Image Process.* **24**(3), 1101–1114 (2015)
26. IP Howard, *Seeing in Depth, Vol. 1: Basic Mechanisms*. (University of Toronto Press, Toronto, 2002)
27. S Birchfield, C Tomasi, A pixel dissimilarity measure that is insensitive to image sampling. *IEEE Trans. Pattern. Anal. Mach. Intell.* **20**(4), 401–406 (1998)
28. R Zabih, J Woodfill, in *Computer Vision—ECCV'94*. Non-parametric local transforms for computing visual correspondence (Springer, New York, 1994), pp. 151–158
29. K Mühlmann, D Maier, J Hesser, R Männer, Calculating dense disparity maps from color stereo images, an efficient implementation. *Int. J. Comput. Vis.* **47**(1–3), 79–88 (2002)
30. C Rhemann, A Hosni, M Bleyer, C Rother, M Gelautz, in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference On*. Fast cost-volume filtering for visual correspondence and beyond (IEEE, Colorado Springs, 2011), pp. 3017–3024
31. CL Zitnick, SB Kang, M Uyttendaele, S Winder, R Szeliski, in *ACM Transactions on Graphics (TOG)*. High-quality video view interpolation using a layered representation, vol. 23 (ACM, Los Angeles, 2004), pp. 600–608
32. J Jiao, R Wang, W Wang, S Dong, Z Wang, W Gao, Local stereo matching with improved matching cost and disparity refinement. *IEEE MultiMedia.* **21**(4), 16–27 (2014)
33. D Scharstein, R Szeliski, Middlebury stereo evaluation-version 2. vision.middlebury.edu/stereo (2002). <http://vision.middlebury.edu/>
34. Z Wang, AC Bovik, HR Sheikh, EP Simoncelli, Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **13**(4), 600–612 (2004)
35. Standard for the quality assessment of three dimensional (3d) displays. 3D Contents and 3D Devices based on Human Factors. *IEEE P3333.1.* **2012** (2012). doi:<http://grouper.ieee.org/groups/3dhf>
36. S Ide, H Yamanoue, M Okui, F Okano, M Bitou, N Terashima, in *Electronic Imaging 2002*. Parallax distribution for ease of viewing in stereoscopic hdtv (International Society for Optics and Photonics, San Jose, 2002), pp. 38–45
37. ITU-R Recommendations, 500.7, methodology for the subjective assessment of the quality of television pictures. Recommendations, ITU-R, Geneva (1995). <http://www.itu.int/rec/R-REC-BT.500-7-199510-S/en>
38. C-C Chang, C-J Lin, Libsvm: A library for support vector machines. *ACM Trans. Intell. Syst. Technol. (TIST).* **2**(3), 27 (2011)
39. M-J Chen, LK Cormack, AC Bovik, No-reference quality assessment of natural stereopairs. *IEEE Trans. Image Process.* **22**(9), 3379–3391 (2013)
40. DL Ruderman, The statistics of natural images. *Netw. Comput. Neural Syst.* **5**(4), 517–548 (1994)
41. PB Hibbard, A statistical model of binocular disparity. *Vis. Cogn.* **15**(2), 149–165 (2007)
42. Y Liu, AC Bovik, LK Cormack, Disparity statistics in natural scenes. *J. Vis.* **8**(11), 19 (2008)
43. A Mittal, AK Moorthy, AC Bovik, No-reference image quality assessment in the spatial domain. *IEEE Trans. Image Process.* **21**(12), 4695–4708 (2012)
44. K Sharifi, A Leon-Garcia, Estimation of shape parameter for generalized gaussian distributions in subband decompositions of video. *IEEE Trans. Circ. Syst. Video Technol.* **5**(1), 52–56 (1995)

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com