Signal Processing 188 (2021) 108179

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Bi-disparity sparse feature learning for 3D visual discomfort prediction ☆

Maryam Karimi [a,*], Mansour Nejati [b], Weisi Lin [c]

[a] Department of Computer Science, Shahrekord University, Shahrekord 88186-34141, Iran
[b] Department of Electrical and Computer Engineering, Isfahan University of Technology, Isfahan 84156-83111, Iran
[c] School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798, Singapore

ARTICLE INFO

Article history:
Received 28 December 2020
Revised 9 May 2021
Accepted 28 May 2021
Available online 2 June 2021

Index Terms:
Stereo image
3D perception
Visual discomfort prediction
Unsupervised feature learning
Sparse representation
Aggregated disparity map

ABSTRACT

Viewing stereoscopic images sometimes causes viewers to feel inconvenience, which is called 3D visual discomfort. Excessive horizontal disparity, misalignment between the left and right views, or depth cues conflicts are some of the important factors involved in 3D visual discomfort. The ability to estimate the degree of 3D visual discomfort can be used to improve the 3D display systems and provide acceptable binocular visual quality. Most of the existing visual discomfort prediction (VDP) approaches extract hand-crafted features based on perceptual modeling and statistical analysis of disparities. We have proposed a simple yet effective VDP model based on unsupervised learning of sparse features which are highly predictive of subjective discomfort levels. These features are extracted from the aggregation of left and right disparity maps. This aggregation effectively highlights the areas with sudden changes and high levels of disparities where discomfort is most likely to occur. The regression model trained by the features, predicts high correlated 3D visual discomfort scores on each dataset. The cross-database results are also superior to other reported ones.

1. Introduction

With the development of various 3D screen types and their growing popularity, 3D content is expected to attract a large portion of the media in the near future. Therefore, quality assurance is of paramount importance in different 3D content services [1]. Quality of Experience (QoE) of 3D visual contents is usually considered to be mainly a combination of the perceived structural visual quality, and visual comfort [2]. The visual quality of stereo images in presence of structural distortions has been extensively studied [3–5] and recently, the visual discomfort that relates to subjective 3D feeling has received much attention. Comparison of the two slightly different images projected to each retina causes the binocular depth perception by human visual system [6]. This perception, called stereopsis, gives the viewer a much broader sense of 3D reality. Despite these advantages, this perception is not always comfortable and pleasant. Physical symptoms such as eye-strain, headache, nausea, fatigue, and focusing difficulty have been reported by many viewers while viewing some 3D images [7–11]. The subjective perception of the difficulties by viewers is called visual discomfort. A variety of physical factors are discovered to be involved in the feeling of visual discomfort including window violation, keynote effects, optical distortions, object compactness, and crosstalk [6–11]. However, the accommodation and vergence mismatches (AVM) are believed to be the most considerable causes of visual discomfort that are related to excessive disparities [12,13]. To deal with the problems of visual quality and safety assurance in 3D services, human subjective assessment is the most reliable solution but it is impossible to use it in real applications due to being expensive and time-consuming. Thus, objective visual discomfort prediction (VDP) metrics are needed for comfortability viewing assessment of stereoscopic images.

In this paper, an unsupervised feature learning method for visual discomfort prediction of stereoscopic images is introduced. We extract both left and right disparity maps and compute an aggregated disparity map for each stereo image pair. These disparities are aggregated to produce a saliency map regarding the image regions with a high potential of causing 3D visual discomfort. A dictionary is then learned on randomly selected patches from the aggregated disparity maps. The sparse coefficients in the sparse representation of each patch demonstrate the contribution of each

☆ This paragraph of the first footnote will contain the date on which you submitted your paper for review. It will also contain support information, including sponsor and financial support acknowledgment. For example, "This work was supported in part by the U.S. Department of Commerce under Grant BS123456."

* Corresponding author.
E-mail addresses: ma.karimi@sku.ac.ir, maryam.karimi.d@gmail.com (M. Karimi), mansour.nejati@ec.iut.ac.ir (M. Nejati), wslin@ntu.edu.sg (W. Lin).

https://doi.org/10.1016/j.sigpro.2021.108179
0165-1684/

dictionary atom. Therefore, we use the sparse representation of local disparities as local visual discomfort descriptors. A weighted average pooling is performed on all sparse vectors to achieve more general features representative of the global visual discomfort level of each stereo pair. This kind of feature learning encodes the disparity information of 3D views especially in neighboring areas with high depth difference usually containing the occluded parts of the scene. A regression model is finally trained on training pooled feature vectors regarding subjective 3D discomfort scores. The performance power of the proposed method is verified by various experimental results on the standard stereo image sets. As mentioned above, we propose a 3D VDP algorithm that, by using simple blocks as well as generating simple features from disparity maps, can provide competitive and even better results than the existing algorithms. In addition, it is more accurate on datasets on which the model is not trained.

In summary our main contributions are: 1) Aggregating both left and right disparity maps to highlight areas containing binocular depth perception information. 2) Unsupervised learning of local visual discomfort aware features using sparse coding which are more discriminative than hand-crafted ones. 3) Composition of local sparse representations by a weighted pooling method proportional to the strength of sparse vectors to provide effective global 3D VDP descriptions.

The remainder of this paper is organized as follows. Section 2 reviews the existing objective VDP methods. In section 3 general aspects of content characteristics that affect visual discomfort are reviewed. A detailed overview of the proposed method is provided in Section 4. In Section 5, after introducing databases and performance measures, we describe experiments to evaluate the performance of the proposed method. Finally, Section 6 concludes the paper.

2. Related work

Objective VDP methods attempt to estimate the discomfort level of stereo images similar to subjective mean opinion scores (MOS). Most of these methods focus on extracting and pooling effective features to be used for training machine learning models to estimate the degree of visual discomfort.

The primary methods presented in this field were based on the statistical features associated with the comfort degrees [14–18]. A statistical study in [14] models the impact of spatial frequency of luminance contrast as well as signed disparity in visual discomfort. A statistical function of luminance magnitude, contrast, spatial frequency, orientation, and disparity was presented in [15] for measuring window violation preference. The relationship between the distribution of parallax and visual comfort was studied by Nojiri *et al.* [16]. A visual fatigue prediction method investigated the effects of excessive vertical and horizontal disparities [18]. Considering the dominant role of excessive disparity, defocus blur and spatial frequency on 3D perception, some common statistical features are extracted from related maps to train a ranking model in [19]. Some other visual comfort statistical features from comfort zone, depth of focus and spatial frequency are integrated to train a robust preference classification model in [20].

Lately, some more advanced methods attempted to exploit the features of the human binocular perception system to estimate 3D visual discomfort. Visual saliency-weighted disparity features were employed in [21]. In this method, a combination of saliency map and disparity gradient was used to extract perceptually and experimentally significant features to extract visual discomfort degrees. Concepts naming local 2D and 3D bandwidths are defined in [22] formulated based on fovea, physiological optics and binocular vision to be used as descriptors to train a VDP model. The VDP model of [23] devised Quantitative models providing features sensitive to accommodation and vergence interactions to predict visual discomfort levels. Additionally, other 3D elements such as sharpness and fusion limits are included in this method. A model-based method extracting coarse and fine statistical and neuronal features from binocular disparities to automatic prediction of 3D discomfort is presented in [24]. Another method showed that object-dependent disparity features such as the mean disparity difference between nearby objects and the object thickness can improve the performance of conventional VDP methods [25]. The authors in [26] attempted to extract feature maps without explicit extraction of disparities. To this end, they use orientation and luminance classification in corresponding patches to find the percentage of unlinked pixels which is descriptive of disparities. Jiang *et al.* extract features based on disparity statistics and neural activities to represent the stereo images in [27]. Then they construct dictionaries from the feature descriptors of images with close MOSs. For each test pair, the reconstruction residuals by each dictionary are used to weight the MOSs in the corresponding group and estimate the final visual comfort score. A sparse representation-based method in [28] used the visual importance disparity and spatial frequency disparity features to learn a dictionary with respect to subjective scores to estimate the 3D discomfort level of stereo images. The authors of [29] in addition to statistical factors of disparity in pixel, focus on angular disparity statistics in three tuned viewing distances, and the neural Middle Temporal Importance (MTI) features [29].

Since visual discomfort is a very complex process in the human visual and perceptual system, it is difficult to determine all its psychological and physical factors and their contribution rates. A fundamental problem with the above methods is the extraction of handcrafted features. Although these features are capable of revealing different perceptual and physiological aspects of 3D visual discomfort, they may miss other phenomena. Additionally, in many of these methods, different disparity map areas are equally important in extracting features, while only certain areas of it lead to visual discomfort. Few methods that try to identify these areas do so depending on the content of the image. Recently, learning of features on raw data has received much attention for various image processing applications as they provide more effective representations than hand-crafted ones [30]. One of the novel methods for learning feature representations is deep learning. A deep learning-based VDP model for stereoscopic images has been presented in [31]. In this method, the spatial right and left features with different receptive fields are hierarchically extracted and fused by a binocular fusion deep network. The encoded features by a disparity regularization network provide disparity relations that embedding them into the deep VDP that improves the final accuracy. In a patch-based convolutional neural network (CNN) model in [30] proxy ground-truth labels are assigned to image patches at first. Then the CNN generates local descriptors that are aggregated by an additional layer into global features. Another method in [32] aggregates features extracted by three separate CNNs to encode 3D perceptual cues and generate final visual comfort scores. The first network extracts visual difference features from left and right views, the second and the third, get saliency weighted absolute disparity and saliency weighted absolute differential disparity maps respectively to extract disparity related feature.

Despite all the benefits, deep-learning approaches have a very time-consuming learning process. In the above methods, the feature learning network and the regression are coupled so the feature learning in such models requires a massive amount of labeled training images and subsequently subjective ratings or some heavy proxy patch labeling step such as the one in [30]. Generally, approaches using hand-crafted features fail to provide robust-enough features, and deep-learning based approaches have a heavy learning phase and may fail to isolate relevant features and predict vi-

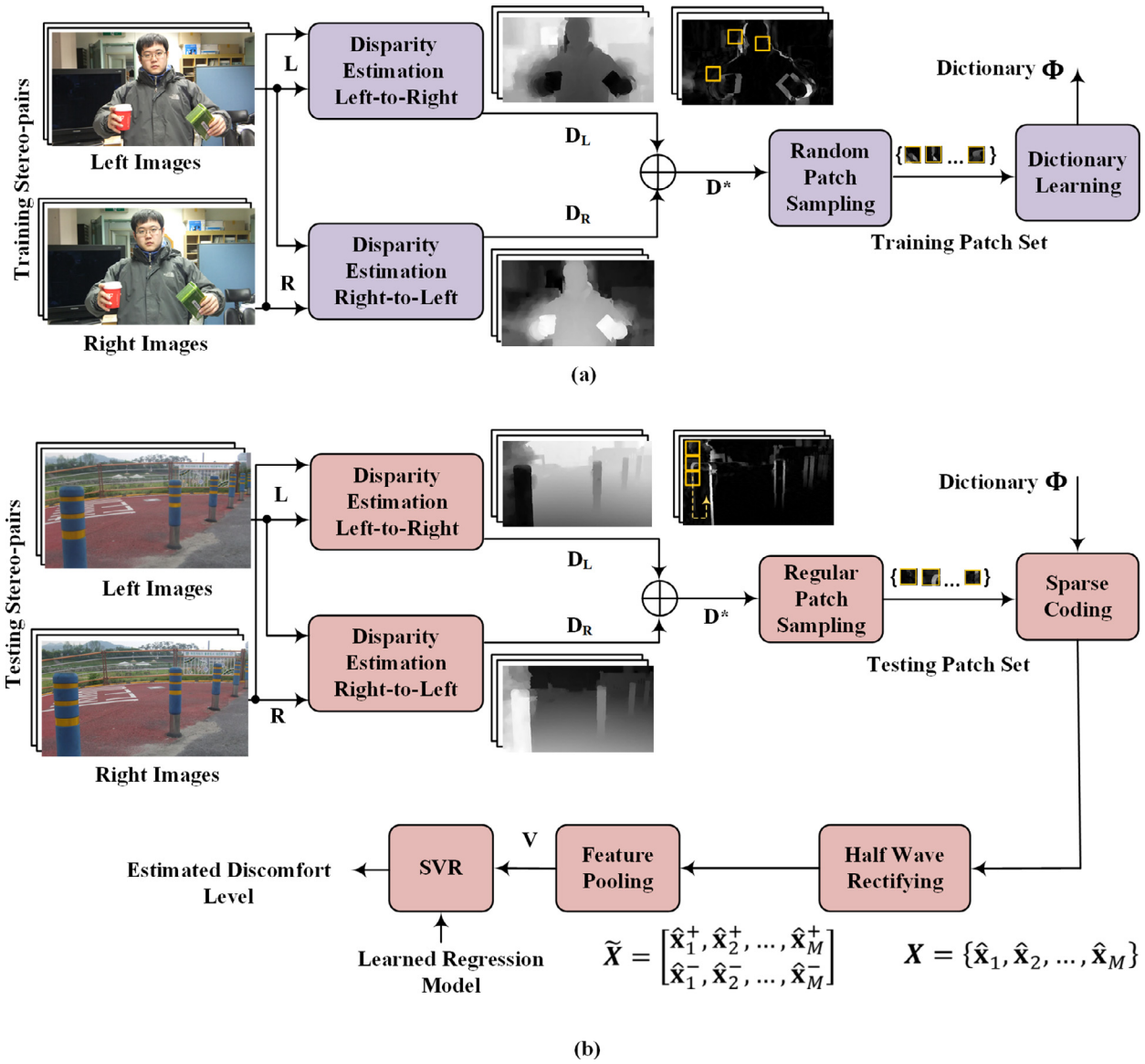

Fig. 1. Diagram of the proposed framework including (a) the unsupervised feature learning phase and (b) the SVR testing phase.

sual discomfort from other types of features that might be content-specific and not 3D-specific.

Unsupervised feature learning algorithms aim to learn suitable representations from unlabeled data as input to a supervised learning machine [33–38]. These methods extract low-dimensional features from high-dimensional data by discovering hidden structures in raw data. Learning sparsifying dictionaries belong to a class of unsupervised feature learning approaches that learn over-complete bases (dictionary elements) for representation of data. The sparse representation of data on such data-adaptive dictionaries can lead to feature vectors that have a higher discriminative power than raw data.

The proposed method in this paper, in addition to providing an unsupervised sparse feature learning method on visual discomfort-aware disparity areas, is easier than deep learning methods since it requires less data for training and with no need for labeling patches. Experiments show that the prediction results obtained by this method are comparable and even better than deep learning-based methods regarding different performance measures. Also, our cross-database test results with different databases indicate that it performs better on datasets that our method is not trained on. Therefore, it is more reliable for any input stereopairs.

3. Proposed method

3.1. Motivation

Physiological studies discovered disparity magnitude and disparity gradient are crucial aspects affecting binocular depth perception [39]. The disparity magnitude refers to the absolute disparity value and the disparity gradient is the disparity difference between adjacent objects.

Disparity Magnitude In general, objects with excessive binocular disparity can lead to diplopia. In fact, when the angular disparity difference between a point and the fixation point is greater than the Panum's fusional limit the human visual system cannot create a single binocular vision [12]. Therefore, the further away the disparity magnitude from this fusional limit, the more visual discomfort may occur [40]. In addition, the depth of focus (DOF) is the retinal defocus degree, which ensures clear vision without changing of accommodation [39]. In 3D display systems the Accommodation and Vergence Conflict (AVC) may occur where the accommodation remains on a 3D screen so that it can receive a clear view, while the vergence changes based on the stereoscopic depth cues to achieve a binocular vision. This conflict which is

directly in relationship with DOF and binocular fusion limit, may cause fatigue and visual discomfort [8]. As a result, the disparity magnitude has been used as the main parameter in designing visual discomfort measures.

Disparity Gradient In addition to the disparity magnitude, the binocular fusion limit is also affected by disparity gradient. Disparity gradient is defined as the disparity difference between nearby objects [25]. In [41] it is reported that when disparity gradient exceeds a certain critical value, binocular fusion can fail even if the disparity magnitude is in the Panum's fusion area. This indicates that the binocular perception comfort related not only to the disparity magnitude of an object but also to the disparity gradient between adjacent objects. Therefore, for objects that have the same disparity magnitude but are adjacent to objects with different disparity differences, they induce different visual discomfort levels.

Since each natural scene contains many objects, and each object may have many adjacent objects, it is necessary to learn features appropriate to the variety of disparity magnitudes and gradients of the scene.

A different discomfort predictor is proposed in this work which incorporates both the disparity magnitude and disparity gradient using the sparse representation of aggregated disparity information as discomfort descriptors. The diagram of the proposed method is depicted in Fig. 1. This method consists of three main steps: unsupervised feature learning, model training, and model testing. In the first step, as shown in Fig. 1(a) two disparity maps, D_L and D_R, are extracted from left to right and right to left views respectively. Then we aggregate them to achieve a map named aggregated disparity (D^*). A set of randomly selected blocks from training D^* maps are finally deployed to learn a sparse coding-based dictionary $\mathbf{\Phi}$. The dictionary learning serves as an unsupervised feature learning approach which exploits the structure underlying the unlabeled data to learn a set of basis features as dictionary elements. In the training step, we first regularly sample non-overlapping patches in each D^* map to be sparse coded using the learned dictionary. Then, the computed sparse features are half-wave rectified to take into account the effect of positive and negative coefficients separately, and pooled to achieve a single global feature vector for each stereo image-pair. A regression model is finally trained using training set vectors. The same steps are performed to extract features from the test set and regression model trained in the previous step is used to predict their discomfort scores as shown in Fig. 1 (b).

3.2. Bi-directional disparity extraction

Recent studies found disparity as very effective information on 3D visual discomfort [18–34]. The feature extraction in the proposed approach is concentrated on disparity information. A common first step to implement a 3D visual discomfort prediction method is some form of disparity estimation, where its accuracy affects the accuracy of the VDP model. To generate the disparity map of each stereo-pair, we chose the optical flow estimation algorithm presented in [42]. The output horizontal pixel displacement map of this algorithm can be referred to as the disparity map of each stereo-pair. We calculate the disparity between the two images, which in our case is the horizontal displacement of a pixel in one of the stereo images with respect to its spatial location in the another one. In our method, disparity is calculated both with respect to the left image L and the right image R to produce two disparity maps D_L and D_R for each stereo image pair. The disparity $D_L(x, y)$ is created such that $L(x, y) = R(x + D_L(x, y), y)$ when matching from left to right. Similarly, $D_R(x, y)$ is obtained by matching from right to left such that $R(x, y) = L(x + D_R(x, y), y)$. If we suppose $[-p, q]$ to be the disparity range in D_L, the disparity values in D_R would be in the range $[-q, p]$. These disparities are

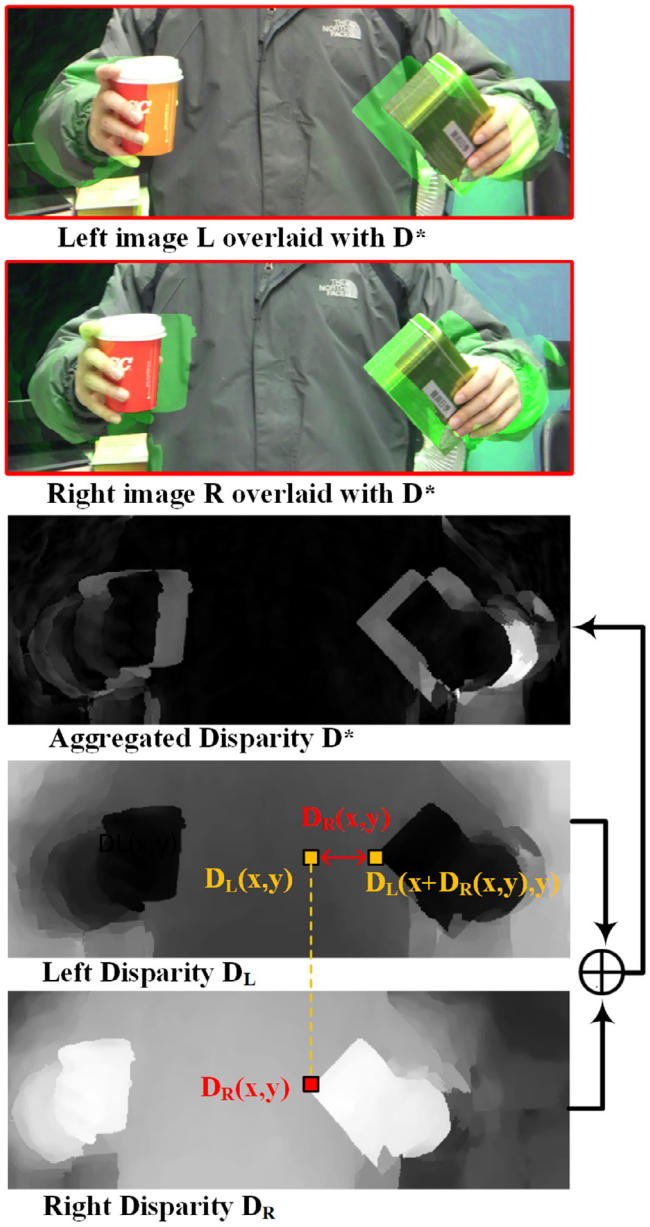

Fig. 2. Left and right images overlaid with aggregated disparity map.

simply aggregated as follows:

$$D^*(x, y) = |D_L(x, y) + D_R(x, y)| \quad (1)$$

Based on the Eq. (1), areas with a large magnitude difference between D_L and D_R will be highlighted in D^*. This usually occurs around the boundaries between foreground and background objects, which usually include occluded areas too as shown in Fig. 2. For more clarification, the aggregated disparity map is displayed as a transparent green layer on the left and right images.

The disparity values are much higher for objects close to the camera (foreground objects) than for objects far from the camera (or background objects). Therefore, in areas near the boundaries between foreground and background objects, one of the views usually contains a background area and the other one contains a foreground.

In such areas, the aggregation of left and right disparities results in larger values than other areas. In other words, for the interior of both the background and foreground objects, the left

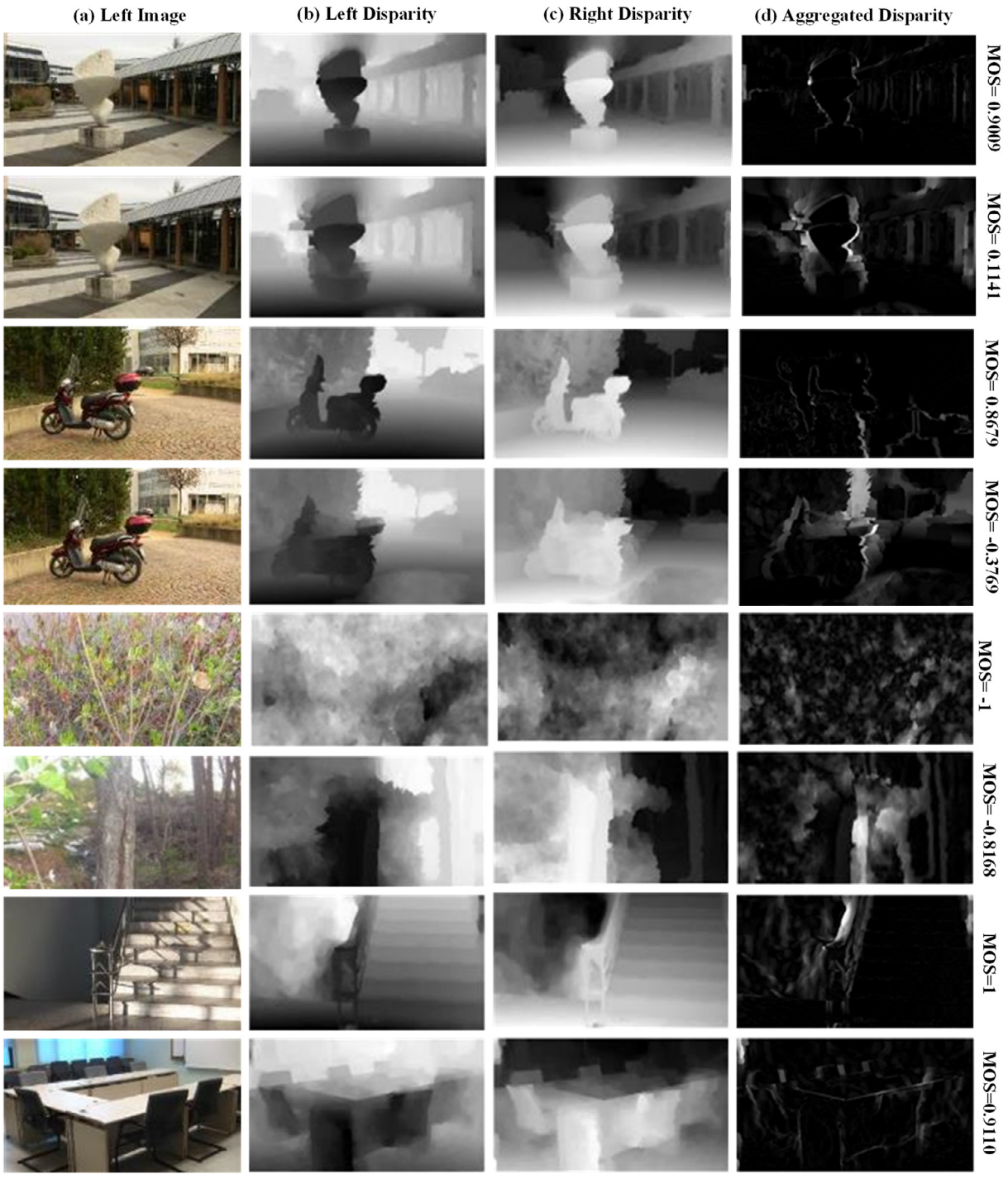

Fig. 3. Eight stereo images with different MOSs, (a) left image, (b) left disparity map, (c) right disparity map, and (d) the generated aggregated disparity maps

and right disparity values almost cancel each other, while in border regions they do not. This aggregation may seem not to work well for high disparity levels in flat regions. However, the disparity differences in the borders with adjacent objects are reflected in the aggregated map. Since the depth changes of the objects give the viewer a sense of 3D perception, these areas appear to have more impact on visual discomfort and other areas without depth changes do not. Accordingly, the absolute summation of the two disparity maps that effectively highlight the sudden and large depth changes such as the occluded regions in 3D perception of stereo images is used to learn descriptors for visual discomfort level estimation. To clarify this fact, Fig. 3 shows the aggregated disparity maps (D^*) generated for eight stereoscopic images with different MOSs. MOS values close to -1 are indicative of high levels of visual discomfort and values close to 1 indicate more visual comfort in 3D perception. Note that some similar images are related to same scenes with different camera distances.

3.3. Dictionary learning

Learning sparsifying dictionaries is an ideal way to be used for local feature encoding. Indeed, dictionary learning based on sparse coding provides us with an unsupervised feature learning

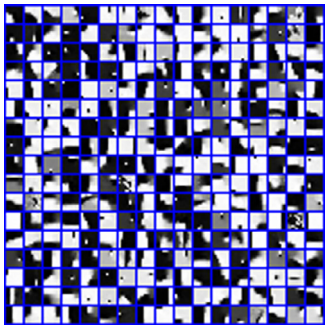

Fig. 4. The dictionary learned on aggregated disparities of IVY LAB dataset.

framework for learning sets of over-complete bases to represent the data space efficiently. Therefore, the sparse representation of data on such data-adaptive dictionaries can lead to feature vectors that have a higher discriminative power than raw data. This unsupervised learning is accomplished by optimizing the dictionary elements (called atoms) for sparse representation of unlabeled input data. The K-SVD algorithm is a successful method in this field which is used in several image processing applications [43]. This method takes a set of N training samples $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N] \in \mathbb{R}^{n\times N}$, and an initial dictionary $\mathbf{\Phi}_0 \in \mathbb{R}^{n\times K}$ $(K > n)$ with K atoms as the inputs. The dictionary is iteratively improved to obtain the sparse representations $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N] \in \mathbb{R}^{K\times N}$ of $\mathbf{Y}$ by solving the following minimization problem.

$$\min_{\mathbf{\Phi},\mathbf{X}} \|\mathbf{Y} - \mathbf{\Phi}\boldsymbol{X}\|_F^2 \text{ s. t. } \forall i \, \|\mathbf{x_i}\|_0 \le L \tag{2}$$

where L is the maximum allowed number of non-zero elements in each sparse code. The $\|\cdot\|_0$ and $\|\cdot\|_F$ denote ℓ_0-pseudo-norm and the Frobenius norm respectively. In each iteration of learning, all the samples in $\mathbf{Y}$ are sparse coded with respect to the current dictionary to produce the sparse matrix $\mathbf{X}$. Then the dictionary atoms are updated by the current matrix **X.**

In our proposed approach, the dictionary is learned on local patches of size $\sqrt{n}\times\sqrt{n}$ pixels which are randomly selected from D^* and rearranged into column vectors. The dictionary is then employed for the encoding of training/testing D^* patches to be used as local discomfort descriptors. One example of the learned dictionary on IVY LAB S3D image dataset [21] is displayed in Fig. 4.

3.4. Sparse representation

Let $\mathbf{Y} = \{\mathbf{y}_i \in \mathbb{R}^n\}_{i=1}^M$ denotes the set of all column-wise non-overlapping $\sqrt{n}\times\sqrt{n}$ patches of a given aggregated disparity map. Solving the following l_1-regularized sparse coding problem results in the sparse representation $\hat{\mathbf{x}}_i \in \mathbb{R}^K$ using the dictionary $\mathbf{\Phi} \in \mathbb{R}^{n\times K}$ for patch $\mathbf{y}_i$:

$$\hat{\mathbf{x}}_i = \underset{\mathbf{x}}{\operatorname{argmin}} \ \|\mathbf{x}\|_1 \ s.t. \ \|\mathbf{y}_i - \mathbf{\Phi}x\|_2^2 \le \lambda \tag{3}$$

where λ is the sparsity regularization parameter and the Least Angle Regression with Lasso modification (LARS) [44] is used to solve this problem. In our method, λ is set to 0.15 and we use the maximum number of steps of the LARS algorithm, L, as a stopping criterion to control the number of non-zero coefficients (i.e. sparsity) of the output solution. The algorithm returns a matrix of sparse vectors $\boldsymbol{X} = \{\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \ldots, \hat{\mathbf{x}}_M\} \in \mathbb{R}^{K\times M}$ for all of the M input patches. These vectors are the sparse representation of data using the most relevant basis elements that are learned on the data itself for which a more discriminative power than raw disparity data has been demonstrated in many image processing and computer vision problems. These local feature vectors are finally pooled together to achieve a global representation for each stereo–pair.

3.5. Feature pooling

Once the sparse coding problem is solved for all non-overlapping patches of an aggregated disparity map D^*, we should integrate all the sparse coefficient vectors to achieve a richer representation of the whole 3D view.

For this purpose, we applied a weighted averaging after a half-wave rectifying to generate a final feature vector.

1) **Half-wave rectifying:** Let $\boldsymbol{X} = \{\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \ldots, \hat{\mathbf{x}}_M\} \in \mathbb{R}^{K\times M}$ be the set of sparse coefficient vectors for a given stereo image pair. Each coefficient vector $\hat{\mathbf{x}}$ is half-wave rectified into two non-negative channels $\hat{\mathbf{x}}^+$ and $\hat{\mathbf{x}}^-$, so that $\hat{\mathbf{x}} = \hat{\mathbf{x}}^+ - \hat{\mathbf{x}}^-$. These are then concatenated to form a new sparse feature vector. In this way, the positive and negative coefficients of the feature vector are attended as two separate vectors. Thus, the effect of both coefficient groups is considered in the final vector. The matrix of half-wave rectified feature vectors $\tilde{\boldsymbol{X}} \in \mathbb{R}^{P\times M}$ is provided by:

$$\tilde{\boldsymbol{X}} = \begin{bmatrix} \hat{\mathbf{x}}_1^+, & \hat{\mathbf{x}}_2^+, & \ldots, & \hat{\mathbf{x}}_M^+ \\ \hat{\mathbf{x}}_1^-, & \hat{\mathbf{x}}_2^-, & \ldots, & \hat{\mathbf{x}}_M^- \end{bmatrix} \tag{4}$$

where $P = 2K$, $\hat{\mathbf{x}}_{i,j}^+ = \begin{cases} \hat{\mathbf{x}}_{i,j} & \hat{\mathbf{x}}_{i,j} > 0 \\ 0 & \hat{\mathbf{x}}_{i,j} \le 0 \end{cases}$ and $\hat{\mathbf{x}}_{i,j}^- = \begin{cases} |\hat{\mathbf{x}}_{i,j}| & \hat{\mathbf{x}}_{i,j} < 0 \\ 0 & \hat{\mathbf{x}}_{i,j} \ge 0 \end{cases}$ in which $\hat{\mathbf{x}}_{i,j}$ denotes the j-th coefficient in the i-th sparse vector $\hat{\mathbf{x}}_i$.

2) **Weighted average pooling:** The matrix $\tilde{\boldsymbol{X}}$ contains a sparse column vector with size $P = 2K$ for each patch. To find a unique feature vector $\mathbf{v} = [v_1, v_2, \ldots, v_P]^T$ describing the discomfort level of the image, we must find a way to integrate the existing columns. To this end after trying several well-known pooling methods we got our best results by using a weighted average (WAVG) pooling which is presented below:

$$v_j = \sum_{i=1}^{\mathrm{M}} w_i \tilde{\mathbf{x}}_{i,j}, \ w_i = \frac{\sum_{j=1}^{\mathrm{P}} \tilde{\mathbf{x}}_{i,j}}{\sum_{i=1}^{\mathrm{M}} \sum_{j=1}^{\mathrm{P}} \tilde{\mathbf{x}}_{i,j}} \tag{5}$$

This weighting let the sparse coefficients in vectors with greater average values have more contribution in the final feature vector **v.** In fact, this method gives more importance to the stronger features in the pooling step.

3.6. Training/testing regression model

After the composition of local sparse descriptors and their integration into the pooled features, it is very important to approximate a function to map the feature space to the MOSs. This function can be used as a discomfort level estimator for testing stereo images. In this work, we train a Support Vector Regression (SVR) model with a Radial Basis Function (RBF) kernel [45]. Suppose $\mathbf{v}_i$ and t_i to be the pooled feature vector and related MOS of i-th stereo image, respectively. Training an SVR model on the training set generates a function $f(\mathbf{v})$ with maximum deviation ε from corresponding MOSs by finding the best values for α and δ:

$$f(\mathbf{v}) = \sum_i \alpha_i \, t_i \exp\left(-\gamma \|\mathbf{v}_i - \mathbf{v}\|^2\right) + \delta \tag{6}$$

where $\mathbf{v_i}$ and $\mathbf{v}$ are two typical training samples and the positive parameter $\boldsymbol{\gamma}$ which controls the radius, is estimated by cross-validation on the training set. In the testing step, the trained function is applied on test vectors to estimate their discomfort scores.

4. Experimental results

4.1. Experimental setup

In both of the dictionary learning and training/testing steps in the proposed method, we extracted 8×8 patches from the images. A dictionary with the size of 300 atoms is learned on 150,000

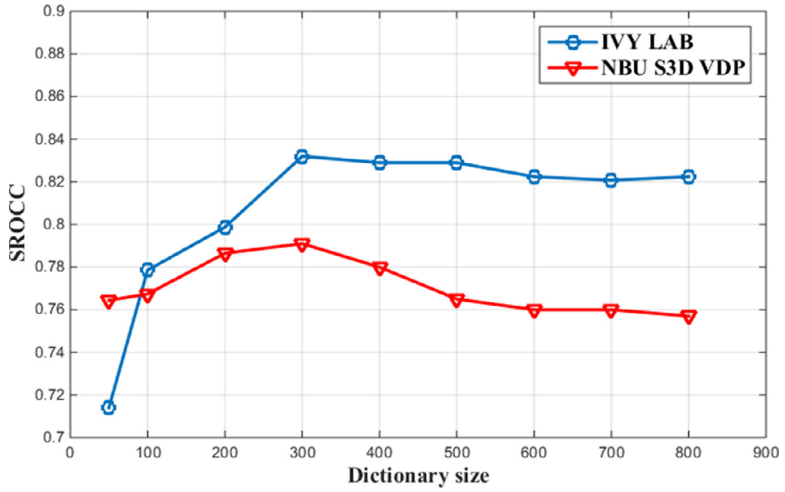

Fig. 5. Accuracy of the proposed method in terms of SROCC with various sizes of dictionaries on IVY LAB and NBU S3D VCA databases.

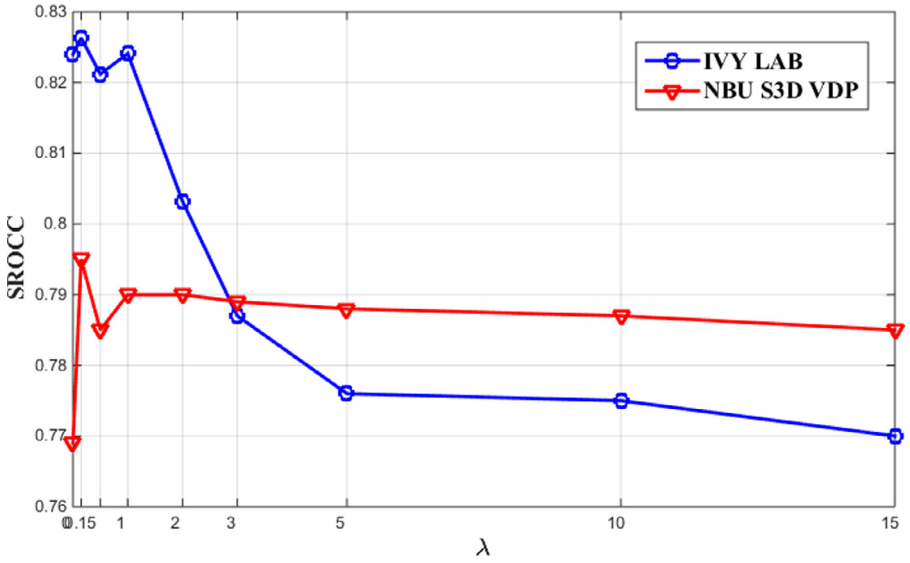

Fig. 6. Accuracy of the proposed method in terms of SROCC with different values of sparsity regularization parameter λ on IVY LAB and NBU S3D VCA databases.

patches from learning aggregated disparity maps in IVY LAB stereo image set based on the optimization problem in (2). In the test step, the vectorized non-overlapping patches of the input images are sparse coded using the learned dictionary according to (3) with sparsity parameter $L = 7$ and the sparsity regularization parameter λ is set to 0.15.

4.2. Datasets

To verify the effectiveness of our proposed method we train and test it on separate parts of IVY LAB [21] and NBU S3D VCA [20] stereo image sets. Also, to validate the generality of the model we perform training and testing the SVR model on distinct datasets. All of the datasets include stereo images with resolution 1920×1080 subjectively assessed by a sufficient number of viewers sitting at a distance three times the height of the display screen and the subjective experiments followed the standard recommendations in ITU-R BT. 500-11 for HDTV [46]. The mean value of different subjective scores for each image is reported as MOS in these datasets. We normalize the MOSs in the range of [-1,1] to have a fair comparison with other existing methods.

IVY LAB consists of 120 indoor and outdoor stereo images with various topics that were taken by a 3D digital camera and the size of the LCD monitor was 40 inches. The MOS values are in the range [1,5], where higher MOSs are indicative of higher visual comfort [21].

NBU S3D VCA has 200 stereo images including 82 outdoor and 118 indoor scenes with a large variety of color, texture and depth ranges which are captured using a Sony HDR-TD30E dual-lens 3D camera [20]. The stereo images were displayed on a 65-inch Ultra HD 3D LED TV and evaluated by 16 subjects with the scores in the range [1,5].

EPFL composed of stereo images with related MOSs in the range [1,100]. Nine different scenes that were captured by cameras with five various distances 10-60 constitute a total of 54 stereo images in this database [47]. A 46-inch polarized stereoscopic display (Hyundai S465D) has been used to display the test stimuli.

4.3. Performance measures

Three well-known performance measures including the Pearson Linear Correlation Coefficient (PLCC), Spearman Rank Order Correlation Coefficient (SROCC), and Root Mean Square Error (RMSE) were used to benchmark the performance of the proposed method against the existing S3D VDP methods. The PLCC and SROCC measure the linearity and monotonicity between the MOSs and pre-

Table 1
Performance comparison of S3D VDP methods on IVY LAB dataset in terms of PLCC, SROCC, and RMSE.

VDP Methods	PLCC	SROCC	RMSE
Nojiri *et al.* [16]	0.703	0.613	0.590
Choi *et al.* [17]	0.682	0.598	0.592
Kim *et al.* [18]	0.711	0.625	0.531
Park *et al.* [22]	0.862	0.781	0.412
Park *et al.* [24]	0.861	0.787	0.413
Jiang *et al.* [27]	0.827	0.818	0.437
Oh *et al.* [23]	**0.865**	0.793	0.405
Oh *et al.* [30]	**0.888**	**0.825**	**0.361**
Proposed	0.829	**0.830**	**0.235**

Table 2
Performance comparison of S3D VDP methods on NBU S3D VCA dataset in terms of PLCC, SROCC, and RMSE.

VDP Methods	PLCC	SROCC	RMSE
Sohn *et al.* [25]	0.787	0.761	0.482
Jung *et al.* [21]	0.778	0.765	0.503
Kim *et al.* [31]	0.813	0.768	**0.402**
Jiang *et al.* [20]	0.808	0.768	0.462
Jiang *et al.* [27]	0.818	0.772	0.456
Jiang *et al.* [19]	**0.835**	0.774	0.589
Yang *et. al.* [29]	**0.866**	**0.803**	0.472
Proposed	0.831	**0.798**	**0.278**

dicted discomfort scores. RMSE measures the prediction consistency. The PLCC is calculated by:

$$PLCC = \frac{cov(X,Y)}{\sigma_X \ . \ \sigma_Y} \tag{7}$$

where X and Y are the subjective and objective quality score sequences, σ_X and σ_Y are their standard deviations, and $cov(X,Y)$ returns the covariance between the two populations. The SROCC applies the PLCC on the corresponding ranks of them, denoted by r_X and r_Y:

$$SROCC = \frac{cov(r_X, r_Y)}{\sigma_{r_X} \ . \ \sigma_{r_Y}} \tag{8}$$

The Root Mean Square Error (RMSE) is a frequently used measure of differences between subjective and objective scores:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i)^2} \tag{9}$$

PLCC and SROCC values closer to 1 are indicative of more accurate estimation while RMSE values close to 0 are ideal for an S3D VDP method. We randomly select 80% of the images from IVY LAB and NBU S3D VCA databases individually as the training sets and the remaining 20% parts are used for testing [31]. Since the number of stereo images in EPFL image set is very limited, like most of the existing related publications, we did not train a model on it and it is used just as a test set to validate the efficiency and the generalizability of the model. We repeat the train/test procedure for 1000 iterations on both the datasets separately and report the median of 1000 values for each of the three measures as the final performance result on each image set regarding that measure.

4.4. Performance evaluation on IVY LAB

To evaluate the performance of the proposed S3D VDP model we report the median PLCC, SROCC, and RMSE on IVY LAB. We compared our performance results with the state-of-the-art VDP methods developed by Nojiri *et al.* [16], Choi *et al.* [17], Kim *et al.* [18], Park *et al.* [22,24], Oh *et al.* [23,30], and Jiang *et al.* [27] in Table. 1. The best two results in each column have been highlighted in bold for a better comparison. It can be seen that the prediction results by the proposed sparse representation based VDP method are better than those of other existing methods in terms of SROCC and RMSE. In the case of the PLCC benchmark, our approach is close and comparable to modern approaches.

Table 3
Results of the performance of cross database validation between IVY LAB and NBU S3D VCA datasets.

Train/Test	NBU/IVY			IVY/NBU		
VDP Methods	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE
Kim *et al.* [18]	0.484	0.365	0.724	0.322	0.281	0.791
Choi *et al.* [17]	0.512	0.372	0.701	0.354	0.307	0.781
Sohn *et al.* [25]	0.552	0.444	0.685	0.458	0.372	0.731
Jung *et al.* [21]	0.698	0.721	0.577	0.645	0.543	0.650
Park *et al.* [24]	0.476	0.561	0.708	0.739	0.559	0.657
Jiang *et. al.* [27]	0.643	0.592	0.624	0.533	0.463	0.693
Yang *et al.* [29]	0.718	0.788	**0.559**	0.671	0.566	0.624
Proposed	**0.721**	**0.796**	0.574	**0.771**	**0.702**	**0.571**

Table 4
Results of the performance of cross database validation training on IVY LAB/NBU S3D VCA and testing on EPFL dataset.

Train/Test	IVY/EPFL			NBU/EPFL		
VDP Methods	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE
Choi *et al.* [17]	0.753	0.801	0.472	-	-	-
Kim *et al.* [18]	0.847	0.850	0.438	-	-	-
Park *et al.* [22]	0.856	0.862	0.420	-	-	-
Park *et al.* [24]	0.853	0.861	0.420	-	-	-
Oh *et al.* [30]	0.859	0.863	0.418	-	-	-
Proposed	**0.923**	**0.930**	**0.287**	**0.907**	**0.918**	**0.435**

4.5. Performance evaluation on NBU S3D VCA

To verify the performance of our proposed method and to ensure that it is not dependent only on a single dataset, we represented aggregated disparity maps of images in NBU S3D VCA dataset [20] by the same dictionary which was learned on IVY LAB dataset. The PLCC, SROCC, and RMSE results on this dataset are compared with related available performance results by different VDP methods in Table 2. The two criteria that deliver the best performance in each column are highlighted in boldface. The results demonstrate the effectiveness of the proposed approach in comparison to others, though the dictionary was trained on another data set. Note that the achieved SROCC and RMSE results are superior to those of the existing schemes.

4.6. Cross-database test

To validate the generality of the S3D VDP model and its dataset less-dependency, we conducted a cross-database test. To this end, we train the model once on IVY LAB and test it on NBU S3D VCA and EPFL [47]. Once again, we train it on NBU S3D VCA dataset and test it on IVY LAB and EPFL image sets. Tables 3 and Table 4 compare the performance of the cross-database tests which are done by different S3D VDP schemes. As shown in the tables the cross-performance is significantly better than all the current methods. Accordingly, the proposed method is not biased to a particular set of data and can be applied to different stereoscopic images.

4.7. Effect of algorithm parameters

The proposed method has several parameters including dictionary size, the sparsity regularization parameter, patch size, the type of feature pooling scheme, and the dictionary learning image

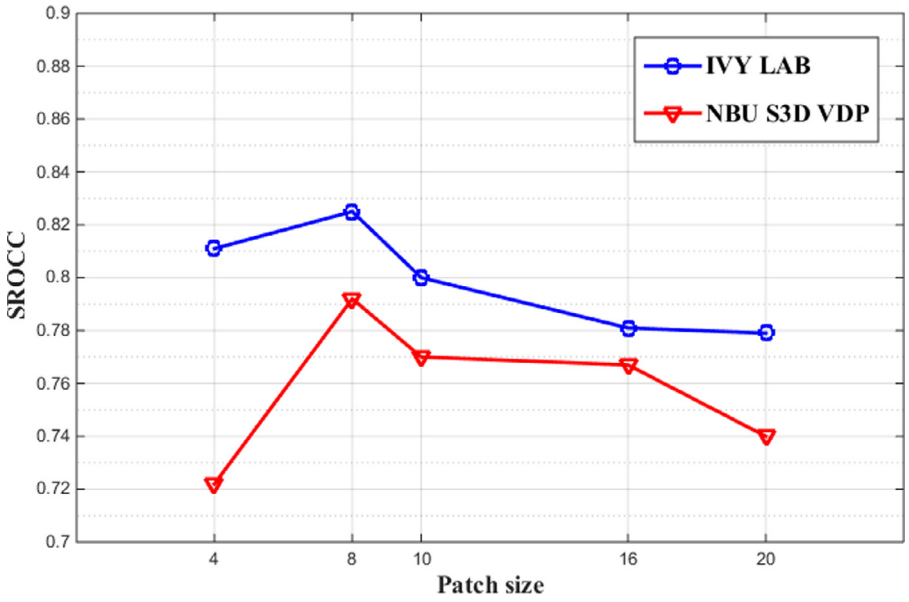

Fig. 7. Accuracy of the proposed method in terms of SROCC with different patch sizes on IVY LAB and NBU S3D VCA databases.

Table 5
Performance of the proposed method in terms of PLCC, SROCC, and RMSE concerning the feature extraction from single and aggregated disparity maps.

Dataset	IVY			NBU		
Map type	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE
Single Disparity	0.672	0.746	0.327	0.780	0.7103	0.302
Aggregated Disparity	**0.829**	**0.830**	**0.235**	**0.831**	**0.798**	**0.278**

Table 6
Performance of the proposed method using different pooling schemes in terms of PLCC, SROCC, and RMSE.

Dataset	IVY			NBU		
Pooling	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE
Max	0.736	0.767	0.287	0.681	0.681	0.374
AVG	0.775	0.804	0.276	0.782	0.743	0.303
SPM	0.748	0.795	0.280	0.676	0.661	0.371
SPA	0.737	0.748	0.291	0.750	0.705	0.326
WAVG	**0.829**	**0.830**	**0.235**	**0.831**	**0.798**	**0.278**

set. The effect of different parameter values on the performance of the model is investigated in this section. Also, the effectiveness of the proposed disparity aggregation method for feature extraction is evaluated.

1) **Dual disparity aggregation**

We did a lot of experimental tests on stereoscopic images with different discomfort levels, and finally, we found that the most related information to the subjective discomfort scores is the disparity. As presented in Section 3.4 we extract our features from the fused disparities obtained in two directions. To demonstrate the impact of using this aggregated map in place of a single disparity map, we retried the dictionary learning, training, and testing phases just on the single disparities obtained from left to right views. The effectiveness of using the mixture of both disparity maps in comparison with a single one is verified in Table 5 for IVY LAB and NBU S3D VCA datasets.

2) **Dictionary size**

The number of dictionary atoms in different applications is set by empirical evaluations. There is a direct linkage between the dictionary size and feature vector length and subsequently the estimation time. Therefore, we learned dictionaries with different sizes and carry out quantitative experiments to find the minimum size of dictionary leading to the best accuracy. For this purpose, the SROCC values between subjective and objective values on each dataset are plotted in Fig. 5 for various dictionary sizes. It can be observed that the performance of the proposed method does not significantly improve if we utilize dictionaries with sizes of more than 300. Accordingly, we chose the dictionary with 300 atoms for the main method.

3) **Sparsity regularization parameter**

The sparse regularization parameter $\lambda \geq 0$ controls the trade-off between data fitting and the sparsity of the solution of the optimization problem in (3). To demonstrate the influence of this parameter values on the performance of our method, we generate solutions with different sparsities by increasing λ from 0 to 15 and measure the SROCC values between subjective and predicted scores on each dataset which are plotted in Fig. 6. It is observed that the best performance results of this method have occurred in $\lambda = 0.15$ for both of the studied datasets.

4) **Patch size**

To analyze the effect of local sampling distance on the performance of the model, five patch sizes including 4×4, 8×8, 10×10, 16×16, and 20×20 were examined for both the dictionary learning and sparse representation steps. In Fig. 5 we plotted the median SROCC of the proposed method according to patch size on IVY LAB and VDP image sets. As shown in Fig. 7 the patch size 8×8 yields a more accurate prediction in our VDP model.

5) **Feature pooling methods**

We tested five different pooling strategies in the proposed framework to fuse the patch sparse representations including max pooling (Max), average pooling (AVG), spatial pyramid max pooling (SPM) [48], spatial pyramid average pooling (SPA) [48], and the proposed weighted average pooling (WAVG) in Section 3.5 The performance on IVY LAB and NBU S3D VCA datasets is listed in Table 6 for each feature pooling method to illustrate the superiority of WAVG comparing to other schemes. According to this table it can be seen that some pooling methods such as MAX and SPM cannot retain and integrate the appropriate sparse coefficients in this

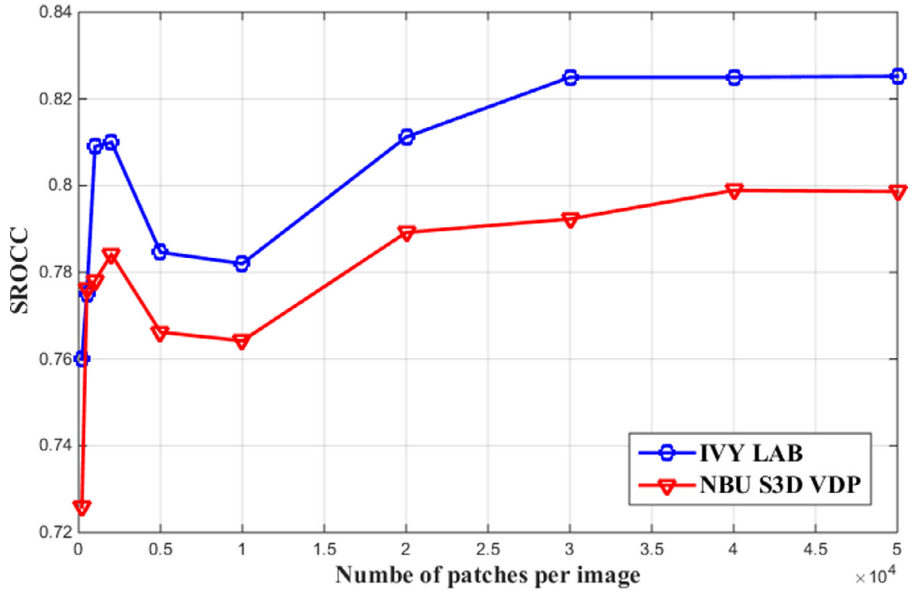

Fig. 8. Accuracy of the proposed method on IVY LAB and NBU S3D VCA databases in terms of SROCC with different numbers of patches selected randomly from each aggregated disparity map of IVY LAB for dictionary learning.

Table 7
Performance of the proposed method on IVY LAB and NBU S3D VCA datasets when the dictionary is learned on each set.

Dictionary	IVY			NBU		
model	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE
IVY	0.822	0.830	0.235	0.801	0.808	0.271
NBU	0.831	0.798	0.278	0.835	0.803	0.266

application. MAX pooling holds only one coefficient in each dimension of the sparse vectors, and therefore the information of other sparse vectors is not included in the final feature vector. Mean pooling method considers the effect of all properties but gives the same weight to all coefficients. The WAVG method considers the effect of all coefficients in proportion to the weight of that feature, which is calculated from the power of sparse vectors. Therefore, it has achieved the best result in our algorithm.

6) Dictionary learning image set

To make sure that the results of this method depend little on the set of images on which the dictionary has been learned, we learned two separate dictionaries on IVY LAB and NBU S3D VCA datasets and performed the training/tests on each set with respect to each of the dictionaries. Although the number of images in each image set is still limited, the very small differences in the results seen in Table 7 show that this method does not depend much on the dictionary learning from a dataset.

7) The number of random patches per image

In order to achieve a proper dictionary based on which the data representation can provide us good descriptors of visual discomfort levels, it is necessary to select a sufficient number of non-flat patches from the aggregated disparity maps for this purpose. To this end we randomly select a number of patches per map on IVY LAB dataset and discard low variance patches from the final set. The effect of this number on performance with both datasets is examined in Fig. 8. The results in this figure demonstrate that selecting more than 40,000 patches from each image will not improve the effectiveness of the dictionary. Accordingly, this number is used in the proposed model.

5. Conclusion

In this paper, we have explored a sparse feature learning method to estimate the 3D visual discomfort level of stereoscopic images. We computed disparity with respect to the left image and the right image for each stereo pair. By aggregating the disparities, we generated a map from which the features are extracted. By learning a sparsifying dictionary, we obtained elementary signals from the data itself that can represent the data space effectively. Therefore, the sparse representation of disparity over a trained data-adaptive dictionary, led to feature vectors that have a higher discriminative power than raw data. The features were finally combined by a weighted pooling strategy to train and test an SVR model for the prediction of 3D visual discomfort levels. Experimental results show that this method achieves better performance results than current methods. The cross-database results also outperform those of all the existing methods.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Maryam Karimi: Conceptualization, Methodology, Software, Validation, Writing - review & editing, Funding acquisition. **Mansour Nejati:** Conceptualization, Methodology, Software, Writing - review & editing. **Weisi Lin:** Writing - review & editing, Supervision.

Acknowledgments

This work has been financially supported by the research deputy of Shahrekord University. The grant number was 99GRN31M49237.

References

[1] B. Mendiburu, 3D Movie Making: Stereoscopic Digital Cinema from Script to Screen, Focal Press, Burlington, MA, USA, 2009.

[2] M. Urvoy, M. Barkowsky, P.L. 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) (2013) 641–655.
[3] F. Shao, W. Lin, S. Gu, G. Jiang, T. Srikanthan, Perceptual full-reference quality assessment of stereoscopic images by considering binocular visual characteristics, IEEE Trans. Image Process. 22 (5) (2013) 1940–1953.
[4] M. Karimi, M. Nejati, S.R. Soroushmehr, S. Samavi, N. Karimi, K. Najarian, Blind stereo quality assessment based on learned features from binocular combined images, IEEE Trans. Multimedia 19 (11) (2017) 2475–2489.
[5] M. Karimi, N. Soltanian, S. Samavi, K. Najarian, N. Karimi, S.R. Soroushmehr, Blind stereo image quality assessment inspired by brain sensory-motor fusion, Digit. Signal Process. 91 (2019) 91–104.
[6] B. Farell, Two-dimensional matches from one-dimensional stimulus components in human stereopsis, Nature 395 (1998) 689–693.
[7] K. Ukai, P.A. Howarth, Visual fatigue caused by viewing stereoscopic motion images: Background, theories, and observations, Displays 29 (2) (2008) 106–116.
[8] M. Lambooij, W.A. IJsselsteijn, I. Heynderickx, Visual discomfort and visual fatigue of stereoscopic displays: a review, J. Imaging Sci. Tech. 53 (3) (2009) 1–14.
[9] W.J. Tam, F. Speranza, S. Yano, K. Shimono, H. Ono, Stereoscopic 3D-TV: visual comfort, IEEE Trans. Broadcast. 57 (2) (2011) 335–346.
[10] Y. Jung, H. Sohn, S. Lee, Y. Ro, Visual comfort improvement in stereoscopic 3D display using perceptually plausible assessment metric of visual comfort, IEEE Trans. Consum. Elec. 60 (1) (2014) 1–9.
[11] T. Bando, A. Iijima, S. Yano, Visual fatigue caused by stereoscopic images and the search for the requirement to prevent them: a review, Displays 33 (2) (2012) 76–83.
[12] D.M. Hoffman, A.R. Girshick, K. Akeley, M.S. Banks, Vergence accommodation conflicts hinder visual performance and cause visual fatigue, J. Vision 8 (3) (2008) 1–30.
[13] T. Kim, S. Lee, A.C. Bovik, Transfer function model of physiological mechanisms underlying temporal visual discomfort experienced when viewing stereoscopic 3D images, IEEE Trans. Image Process. 24 (11) (2015) 4335–4347.
[14] S.P. Du, D. Masia, S.M. Hu, D. Gutierrez, A metric of visual comfort for stereoscopic motion, ACM Trans. Graphics (TOG) 32 (6) (2013) 222:1-222:9.
[15] S. Poulakos, R. Monroy, T. Aydin, O. Wang, A computational model for perception of stereoscopic window violations, in: IEEE Int'l Wkshp Quality of Multimedia Experience (QoMEX), 2015, pp. 1–6.
[16] Y. Nojiri, H. Yamanoue, A. Hanazato, F. Okano, Measurement of parallax distribution and its application to the analysis of visual comfort for stereoscopic HDTV, in: SPIE, Stereoscopic Displays Virtual Reality Syst. X, 5006, 2003, pp. 195–205.
[17] J. Choi, D. Kim, S. Choi, K. Sohn, Visual fatigue modeling and analysis for stereoscopic video, Opt. Eng. 51 (1) (Jan. 2010) 017206.017206-11.
[18] D. Kim, K. Sohn, Visual fatigue prediction for steroscopic image, IEEE Trans. Circuits Syst. Video Technol. 21 (2) (2011) 231–236.
[19] Q. Jiang, F. Shao, W. Lin, G. Jiang, On predicting visual comfort of stereoscopic images: a learning to rank based approach, IEEE Signal Process Lett. 23 (2) (2016) 302–306.
[20] Q. Jiang, F. Shao, G. Jiang, M. Yu, Z. Peng, Three-dimensional visual comfort assessment via preference learning, J. Electron. Imaging 24 (4) (2015) 043002.
[21] Y. Jung, H. Sohn, S. Lee, H. Park, Y. Ro, Predicting visual discomfort of stereoscopic images using human attention model, IEEE Trans. Circuits Syst. Video Technol. 23 (12) (2013) 2077–2082.
[22] J. Park, S. Lee, A.C. Bovik, 3D visual discomfort prediction: vergence, foveation, and the physiological optics of accommodation, IEEE J. Select. Topics Signal Process. 8 (3) (2014) 415–427.
[23] H. Oh, S. Lee, A.C. Bovik, Stereoscopic 3D visual discomfort prediction: a dynamic accommodation and vergence interaction model, IEEE Trans. Image Process. 25 (2) (2016) 615–629.
[24] J. Park, H. Oh, S. Lee, A.C. Bovik, 3D visual discomfort predictor: Analysis of disparity and neural activity statistics, IEEE Trans. Image Process. 24 (3) (2015) 1101–1114.
[25] H. Sohn, Y. Jung, S. Lee, Y. Ro, Predicting visual discomfort using object size and disparity information in stereoscopic images, IEEE Trans. Broadcast. 59 (1) (2013) 28–37.
[26] J. Chen, J. Zhou, J. Sun, A.C. Bovik, Visual discomfort prediction on stereoscopic 3D images without explicit disparities, Signal Process. Image Commun. 51 (2017) 50–60.
[27] Q. Jiang, F. Shao, G. Jiang, M. Yu, Z. Peng, Visual comfort assessment for stereoscopic images based on sparse coding with multi-scale dictionaries, Neurocomputing 252 (2017) 77–86.
[28] H. Xu, G. Jiang, M. Yu, T. Luo, Z. Peng, F. Shao, H. Jiang, 3D visual discomfort predictor based on subjective perceived-constraint sparse representation in 3D display system, Future Genera. Comput. Syst. 83 (2018) 85–94.
[29] J. Yang, V. Nguyen, K. Sim, Y. Zhao, W. Lu, 3-D visual discomfort assessment considering optical and neural attention models, IEEE Trans. Broadcast (2019).
[30] H. Oh, S. Ahn, S. Lee, A.C. Bovik, Deep visual discomfort predictor for stereoscopic 3d images, IEEE Trans. Image Process. 27 (11) (2018) 5420–5432.
[31] H.G. Kim, H. Jeong, H.T Lim, Y.M. Ro, Binocular fusion net: deep learning visual comfort assessment for stereoscopic 3D, IEEE Trans. Circuits Syst. Video Technol 29 (4) (2018) 956–967.
[32] H. Jeong, H.G. Kim, Y.M. Ro, Visual comfort assessment of stereoscopic images using deep visual and disparity features based on human attention, in: 2017 IEEE International Conference on Image Processing (ICIP), 2017, pp. 715–719.
[33] P. Ye, J. Kumar, L. Kang, D. Doermann, Unsupervised feature learning framework for no-reference image quality assessment, in: Proceeding IEEE Conference Computer Visual Pattern Recog, 2012, pp. 1098–1105.
[34] Z. Zhang, F. Li, M. Zhao, L. Zhang, S. Yan, Joint low-rank and sparse principal feature coding for enhanced robust representation and visual classification, IEEE Trans. Image Process. 25 (6) (2016) 2429–2443.
[35] Z. Li, Z. Zhang, J. Qin, Z. Zhang, L. Shao, Discriminative fisher embedding dictionary learning algorithm for object recognition, IEEE Trans. Neural Netw. Learn. Syst. 31 (3) (2019) 786–800.
[36] Z. Zhang, W. Jiang, J. Qin, L. Zhang, F. Li, M. Zhang, S. Yan, Jointly learning structured analysis discriminative dictionary and analysis multiclass classifier, IEEE Trans. Neural Netw. Learn. Syst. 29 (8) (2017) 3798–3814.
[37] Q. Ye, L. Fu, Z. Zhang, H. Zhao, M. Naiem, Lp-and Ls-norm distance based robust linear discriminant analysis, Neural Netw. 105 (2018) 393–404.
[38] Z. Zhang, S. Yan, M. Zhao, Pairwise sparsity preserving embedding for unsupervised subspace learning and classification, IEEE Trans. Image Process. 22 (12) (2013) 4640–4651.
[39] I.P. Howard, in: *"Seeing in depth," Vol. 1: Basic mechanisms*, University of Toronto Press, 2002, pp. 272–273.
[40] S. Yano, M. Emoto, T. Mitsuhashi, Two factors in visual fatigue caused by stereoscopic HDTV images, Displays 25 (4) (2004) 141–150.
[41] P. Burt, B. Julesz, A disparity gradient limit for binocular fusion, Science 208 (4444) (1980) 615–617.
[42] D. Sun, S. Roth, M.J. Black, Secrets of optical flow estimation and their principles, in: 2010 IEEE Computer Society Conference Computer Vision Pattern Recognition, 2010, pp. 2432–2439.
[43] M. Aharon, M. Elad, A. Bruckstein, K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation, IEEE Trans. Signal Process. 54 (11) (2006) 4311–4322.
[44] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, Least angle regression, Ann. Stat. 32 (2) (2004) 407–499.
[45] B. Scholkopf, A.J. Smola, Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond, MIT Press, Cambridge, MA, USA, 2002.
[46] Recommendation, ITU-R BT. 500-11, Methodology for the subjective assessment of quality of television pictures, Stand. Sect. ITU (2002).
[47] L. Goldmann, F.D. Simone, T. Ebrahimi, Impact of acquisition distortions on the quality of stereoscopic images, International Workshop Video Process. Quality Metrics Consumer Elec. (VPQM), 2010.
[48] S. Lazebnik, C. Schmid, J. Ponce, Beyond bags of features: spatial pyramid matching for recognizing natural scene categories, Proceeding IEEE Computer Society Conference Computer Visual Pattern Recoginition 2 (2006) 2169–2178.