Robustness of Reflection Symmetry Detection Methods on Visual Stresses in Human Perception Perspective

IBRAGIM R. ATADJANOV, SEUNGKYU LEE
Department of Computer Science and Engineering, Kyung Hee University, 1732, Deogyeong-daero, Giheung-gu, Yongin-si, Gyeonggi-do, 17104, Republic of Korea (e-mail: ibragim@khu.ac.kr, seungkyu@khu.ac.kr)
Corresponding author: Seungkyu Lee (e-mail: seungkyu@khu.ac.kr).

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2018-2013-1-00717) supervised by the IITP (Institute for Information & communications Technology Promotion).

ABSTRACT  Symmetry is one of the most frequently observed fundamental regularities in the visual characteristic of real-world objects. The human brain has been trained to respond quickly to symmetry patterns, organizing them as salient clues for the unique description of objects. Recently, automatic symmetry detection methods have been widely introduced in computer vision and graphics fields for 2-dimensional and 3-dimensional object data including reflection, translation, and rotation symmetry patterns. Researchers have invented features inspired by a human vision system and have adopted deep learning approaches. On the other side, traditional performance evaluations have been conducted on a unified test dataset containing random degrees of diverse visual challenges. However, they ignore observing the insight of usability and practicality of the methods in higher level tasks such as object recognition. In this work, we carefully organize visual stress dataset for reflection symmetry detection evaluation proposing a novel evaluation framework. The state-of-the-art reflection symmetry detection methods are re-evaluated and analyzed in human perception perspective.

INDEX TERMS  reflection symmetry, performance evaluation, human symmetry perception, visual stresses, psychophysics

I. INTRODUCTION
Symmetry is ubiquitously observed phenomena in both natural [1] and human-made objects [2], [3]. Many living organisms such as birds [4], animals [5] and insects [6]–[8] perceive symmetry patterns of the natural environment. It has been studied that there exists a significant correlation between symmetry and aesthetics, excellence in manufacturing and health [4], [5], [9]–[11]. Symmetry perception is also the matter of survival for jungle animals helping them recognize the natural enemy. Because symmetry is fundamentally inherent to objects of this world, symmetry perception is evolved as a crucial necessity for visual object recognition for humans as well [12]. It is determined that human can perform core object recognition task in a fraction of seconds [13]. The human object recognition process, which is believed to be performed in the ventral visual stream of the brain cortex, consists of multiple stages: line and edge detection [14]–[17], shape representation (e.g. grouping the stimuli coming from retina’s visual sensors) [18]–[21] and symmetry perception [22], [23]. Being in complex interaction with object recognition in the human brain, symmetry perception is a preattentive process.

Due to the critical role of symmetry patterns in perceiving and understanding our world, automatic symmetry detection has been widespread interest in many research fields such as neuroscience [24], psychology, and computer science [2]. For the last few decades, researchers introduced numerous computational symmetry detection methods [25], [26]. However, symmetry detection from real-world images is still a challenging task in computer vision and pattern recognition. There have been several symmetry detection competitions [26]–[28] on public evaluation dataset for diverse symmetry types. Our observation on such evaluations of symmetry detection methods is that traditional evaluation scheme calculates a score and ranking based on a single fixed public dataset, but it does not provide detailed insight into strength
A novel evaluation framework for computational reflection symmetry detection methods is built upon the knowledge in two emerging directions of symmetry study: (1) human symmetry perception, and (2) computational symmetry detection methods.

A. HUMAN SYMMETRY PERCEPTION

The origin of human symmetry perception studies dates far back to XIX century [31]. Since then, the numerous concepts of modeling human symmetry perception process are proposed and empirically validated by conducting various psychophysical experiments [3]. In order to understand internal brain processes responsible for symmetry perception, researchers of neuroscience [32] and psychophysics [23] fields investigated human brain activities using fMRI (Functional Magnetic Resonance Imaging). They observed a correlation between symmetry perception and activations in V3A, V4, V7 and LO, DLO regions of brain cortex [22], [33].

Regarding work of Tyler et al. [34], the symmetry detection is three stages of processes: (1) elaboration of dimensionality of stimuli properties and passing the information to neural analyzers that impose varieties of symmetries; (2) the self-matching feed-forward process that is performed in parallel on each feature across all possible symmetries; (3) the active and manipulative recognition process, which identifies object properties that are too complex to perform at previous stages. Despite extensive experiments and researches dedicated to understanding the mechanics and physics of human symmetry perception, it is still unclear how the human brain perceives symmetry such effortlessly, and what exact neural processes are responsible for it.

II. BACKGROUND AND RELATED WORK

The proposed evaluation framework for computational reflection symmetry detection methods is based on the human organized dataset and demonstrate behavior and robustness of computational methods on each visual stress type.

FIGURE 1: The workflow of the proposed work. (a) Provided images for each visual stress type, (b) determine the human perception threshold for each visual stress type and (c) generate a human organized dataset for each stress type. (d) Evaluate computational symmetry detection methods on the human organized dataset and demonstrate behavior and robustness of computational methods on each visual stress type.
Among symmetries, reflection symmetry is more salient type [35] for humans. Researchers vastly studied reflection symmetry and psychophysically revealed the properties that attract human perception [35]–[50]. Regions supporting the symmetry structure are called the integration regions. The perturbations and distortions in those integration regions are perceived much easily [35]–[38] by humans. The shape and size of the integration region of reflection symmetry pattern scale along with its spatial frequencies [39]. Authors psychophysically determined that the integration region is 2:1 aspect ratio radii ellipse where longer radius equals the length of the reflection line. As eccentricity increase in human eye retina, the integration region of symmetry gets narrower [40]. In other words, the symmetry pattern that falls on the peripheral vision has a narrow integration region. Therefore, symmetry detection is found preattentive only in fovea area of the retina [41]. However, it is still possible to discriminate and detect symmetry in the periphery area of the retina, but that symmetry pattern has small perceptual strength [38], [42].

The orientation of the symmetry pattern has a little effect on symmetry detection [43], [44]. However, the orientations of supporting features in the integration region of the symmetry pattern have a considerable impact to the robustness of symmetry perception; humans have the higher robustness to features orthogonally oriented to symmetry axis than to those that have parallely oriented [45]. Human symmetry perception is robust to distortions caused by perspective projection too. Moreover, the perception of skewed symmetries helps to perceive the orientation of 3D surfaces which contain those skewed symmetries [46]. Therefore, the perception of skewed symmetries is a crucial tool in the judgment of object orientations in space [47]. Another interesting outcome of psychophysical experiments shows that having the opposite contrast on reflected feature pairs makes the symmetry pattern imperceptible [48]–[50].

In order to adopt the properties of human symmetry perception into computational symmetry detection methods, researchers introduced a goodness measure for symmetry patterns. Symmetries with high goodness are detected easily by a human. Van Der Helm et al. introduced a holographic model which measures the goodness of symmetry pattern as a ratio between the number of features supporting the symmetry and the number of all features [51], [52]. The model does not care about the qualitative properties of features and quantitatively measures the goodness. Wageman et al. proposed a qualitative model - the bootstrap model - which exploits the features’ orientations and locations in the visual field, perturbation information, and groupings as summarized in [53]. Later, with the initiative of Wagemans et al. [54], van Der Helm et al. showed how his quantitative goodness model could be combined with qualitative bootstrap model [55]–[57]. Dakin and his colleagues also introduced qualitative (process) models that consider eye fixation information, the orientation of the features, size and shape of the integration region [39], [41], [58], [59].

B. COMPUTATIONAL SYMMETRY DETECTION METHODS AND EVALUATION

Reflection symmetry is one of the most occurred regularities, and it has considerable insight into object recognition and scene understanding. Many researchers were devoted to finding practical and robust computational reflection symmetry detection solutions for a few decades as summarized in [2]. A decade ago, the authors pointed out challenges that are needed to be addressed in order to utilize the favor of symmetry for artificial intelligence. Since then, various approaches were proposed and evaluated in competition workshops of high-level conferences like [26]–[28]. Lately, shape and structure information of an object gained more interest for symmetry detection. Methods utilizing this information for symmetry detection have shown improvements in challenging natural images [60]–[62]. Latest appearance (color, texture) [62], [63] and patch-kernel based methods [64]–[66] also, achieved the state-of-the-art performance by exploiting symmetry detection task as energy minimization [63], registration [67] and linear assignment problem [30]. Despite the interest in, and the number of works on symmetry detection has been increasing year by year, the state-of-the-art method of last decade still competitive with current ones. Based on last symmetry detection competition workshop [26], the method proposed in [68] even outperformed the current state-of-the-art on multiple symmetries detection in 2D real-world dataset.

The evaluation method proposed in symmetry detection competition workshop [27] is accepted and being used as a standard evaluation for most of the computational symmetry detection papers proposed since. As mentioned by researchers multiple times, the analytical evaluation of the symmetry is a nontrivial task. The rules for determining the correct detection is decided empirically and not always achieved a fair evaluation. In [61], Atadjanov and Lee demonstrated one of the weak sides of the traditional evaluation method which evaluates the performance having a limited amount of ground truth information. They demonstrated that the ground truth set for multiple symmetries datasets does not include all potential symmetries. Considering the detection of provided ground truth symmetries as true positives is not a fair evaluation for a method that can detect other unlabeled potential symmetries. Therefore, the authors additionally evaluated the performance by judging each detected axis by human evaluators.

We have encountered different kinds of method evaluations in the recent state-of-the-art symmetry works. Authors in [29] provided the evaluation that shows the behavior of their symmetry map generation method. They evaluated reflection correspondence rate based on the distance measure between estimated reflection correspondence points and ground truth correspondence. They showed that their method achieves more improved symmetry map as the number of iterations increase in their method. By this, they proved the convergence of the proposed iteration based method. The evaluation they provide shows the behavior of their method.
over increasing iterations which does not match with our purpose and that evaluation is only specific to their proposed method algorithm, which is iteration based. In [30], the authors provide reflection symmetry plane detection for 3D cloud points. In order to show the robustness of their detection method for visual distortions, they evaluate their method by applying perturbations to the cloud data with different perturbation variances. To our point of view, this method is useful to show the performance as a behavioral function of perturbations. However, the method is not mature and need to be generalized in order to be applied universally for all possible symmetry works. For example, they applied the perturbation to the 3D point cloud, which is not possible for 2D real-world images. Our proposed framework demonstrates behavior for more broader, specific visual stresses which more likely covers real-world scenarios and uses human symmetry perception as a reference.

C. SIMILAR WORKS IN OTHER FIELDS

Recently, researchers of object recognition field proposed a comparative evaluation of deep neural networks against humans on images under various visual stresses [69], [70]. In [69], the authors use four visual stress types: color, contrast, noise, and eidolon. For the color experiment, the authors compare detection rates on two, color and grayscale, conditions. For each noise and contrast stress type, the authors use fixed eight stress levels. In eidolon-experiment, 24 different conditions were employed. In [70], the authors utilized blur and noise visual stress type, with only five different stress levels (standard deviation). In the experiment, participants are asked to categorize the presenting image into a fixed number of categories. All work showed that as the intensity of stress increase, human detection achieves better performance than computational methods. To our knowledge, the experimental method that is utilized in [70] is inadequate. The method presents images with descending stress level and stops at the stress level where a participant successfully recognizes the stressed image. So, an accidental miss-response from a participant makes the result of the experiment invalid. Nevertheless, in both works above, the intensities of the visual stress seem to be sampled by the power law (or Weber-Fechner law). However, we doubt that human symmetry perception and visual stress intensity follows Weber-Fechner law. For example, in [71], authors reported that detection of symmetry in the presence of noise does not follow Weber-Fechner law, but it follows the psychophysical law that holds for glass pattern. Moreover, our proposed work utilizes multiple other visual stress types, and we cannot know the underlined relationship between the human symmetry perception and each of those visual stress types. Therefore, we adopt the up-down staircase method [72] with variable step-size in order to determine the absolute threshold of each visual stress type carefully. The threshold indicates human symmetry perception limit. Please, refer to the literature [73] for detailed information about psychophysics and psychophysical methods.

III. VISUAL STRESS DATASET IN HUMAN SYMMETRY PERCEPTION PERSPECTIVE

Human symmetry perception is robust to various visual stresses. We define 11 visual stress types on reflection symmetry patterns. We psychophysically find the absolute threshold for each visual stress type. The absolute threshold is the biggest intensity level of visual stress (the smallest intensity level of stimulus) at which a human still can perceive the symmetry. Next, we conduct two psychophysical experiments. The first experiment determines the absolute threshold for each image of all visual stress types. The second experiment determines the absolute threshold for each visual stress type.

A. VISUAL STRESS TYPES

In this work, we define five primitive types of visual stresses: (1) blurring, (2) brightness, (3) additive white noise, (4) size/resolution, and (5) affine skewness. For utilizing the psychophysical method, the relationship between visual stress type and human symmetry perception (sensation) strength should be monotonic; the more significant the amount of visual stress in an image, the less the human symmetry perception strength. This requirement applies to all types of visual stresses except brightness, because, when an image is at perceptually optimal brightness, both the increase and the decrease of brightness cause visual stress. Therefore we divide brightness stress into two types: positive brightness change (brightening) and negative brightness change (darkening). We apply blurring, brightening, darkening and additive white noise in two ways: stress in whole image and stress in one half of reflection symmetry pattern. Stress in one half simulates reflection symmetry pattern on a mirror-like surface such as water surface. Followings are eleven visual stress types evaluated in this work:

- **Blur Half (BH).** Gaussian blurring one half of symmetry pattern in an image. The standard deviation of Gaussian blurring is used to change the strength of visual stress.
- **Blur Whole (BW).** Gaussian blurring the whole image. The standard deviation is used to change the intensity of visual stress.
- **Brightness Half (LH).** Increasing the brightness in one half of symmetry pattern in an image. The intensity of image colors defines the intensity of visual strength.
- **Brightness Whole (LW).** Increasing the brightness of the whole image. The intensity of image colors define the intensity of visual strength.
- **Darkness Half (DH).** Darkening one half of symmetry pattern in an image. A decrease in the intensity of image color increases the intensity of visual stress.
- **Darkness Whole (DW).** Darkening whole image. A decrease in the intensity of image color increases the intensity of visual stress.
- **Noise Half (NH).** Adding white noise to one half of symmetry pattern in an image. The standard deviation of white noise is used to change the intensity of visual stress.
Ibragim Atadjanov and Seungkyu Lee: Preparation of Papers for IEEE TRANSACTIONS and JOURNALS

B. PSYCHOPHYSICAL METHOD

As a psychophysical method, up-down staircases [74] with variable step-size is used. Starting intensity $I_s$ is selected randomly using the following formula.

$$I_s = (b_1 + (b_2 - b_1) \cdot \text{rand()} \cdot I_{\text{max}},$$

where $I_{\text{max}}$ is upper bound of the visual stress intensity and $0 \leq b_1 < b_2 \leq 1$. Note that lower bound of visual stress intensity is zero. $\text{rand()}$ is a function that generates random real number between 0 and 1. Step-size $\Delta I$ is updated when three consecutive reversals occur. The following is the update rule formulation.

$$\Delta I = (l_1 + (l_2 - l_1) \cdot \text{rand()} \cdot \Delta I,$$

where $0 \leq l_1 < l_2 \leq 1$ and defines the constraints of the step-size for the update formulation. Above formulation updates the step-size by a random factor between $l_1$ and $l_2$. For each threshold detection procedure, two interleaved staircases are used (see Fig. 3). The first starts with high stress intensity $I_{s1} = (b_1 + (1 - b_1) \cdot \text{rand()} \cdot I_{\text{max}}$; second starts with low stress intensity $I_{s2} = (b_2 - 0) \cdot \text{rand()} \cdot I_{\text{max}}$. In order to decrease the probability of participants guessing the presenting trials content and structure, the visual stress type, the image of that stress type and one of two staircase processes of that image are randomly selected and presented to the participant.

**B. PSYCHOPHYSICAL METHOD**

A visually stressed symmetry image is presented to a participant with the symmetry axis drawn on it. A participant is asked if the image contains a reflection symmetry pattern by provided symmetry axis. The symmetry axis drawn on the image gets the participant’s eye fixated on the reflection symmetry pattern of interest. In this work, we are not interested in reaction time or detection time to the symmetry pattern. Therefore, there is no limitation for both stimulus presentation and the participant’s response time.

**FIGURE 2:** Dataset of images for all 11 visual stress type. Columns represent visual stress types and labeled with visual stress labels. Images are taken from ICCV 2017 symmetry workshop [26].

- **Noise Whole (NW).** Adding white noise to the whole image. The standard deviation of white noise is used to change the intensity of visual stress.
- **Size/Resolution (R).** Changing the size/resolution of an image by a factor. The factor parameter (0 and 1) indicating the change in size is used to change the intensity of visual stress. E.g., 0 means that no change in the size, 0.5 means that the size is halved, 1 means that the size is 0.
- **Skew Across Symmetry Line (SC).** 3D rotating the image plane across the reflection symmetry axis. The rotation angle is used to change the intensity of visual stress.
- **Skew Along Symmetry Line (SA).** 3D rotating the image plane along the reflection symmetry axis. The rotation angle is used to change the intensity of visual stress.

Fig. 2 presents dataset images for all 11 visual stress types.
FIGURE 3: Two interleaved up-down staircases. Sample convergence of threshold is illustrated for two interleaved threshold determination processes.

FIGURE 4: A screenshot from the web page describing sample stressed images. Stressed images are presented for building up the perceptual space in the human brain.

FIGURE 5: A screenshot of the experiment page. A stressed image is presented, and a participant’s response is collected. The progress of experiment completing is also displayed.

C. WEB BASED UI AND EXPERIMENTAL SETUP

We built a web-based system (tool) in order to conduct the psychophysical experiment for multiple participants simultaneously. Participants are requested to register and attend the experiment. Before the experiment starts, the system provides the participants with the definition of the reflection symmetry with multiple example pictures. Participants are also provided the instruction about the experiment and the user interface. Next, the web-based system provides multiple samples of stressed images for each visual stress type. It is done to help the human brain to develop a perceptual scale and measure for each visual stress. Fig. 4 illustrates a screenshot from the web-based system for sample images with LH stress type. After exploring all samples with various stress level from all visual stress types, the actual experiment starts. On the experiment page, a participant is requested to select one of two choices: "symmetry" or "not symmetry" for each presenting stressed image. The user interface also provides the tentative experiment progress. Fig. 5 shows the screenshot of the experiment page.

D. HUMAN EXPERIMENTS

Experiment 1: Absolute Threshold for Each Image of Visual Stress Types

In this experiment, for all 55 images (5 images for each visual stress type), two interleaved processes are created. A process describes one up-down staircase with at least 20 trials for each. At each presentation, one visual stress type is selected randomly among eleven. Afterward, one image from the dataset of that stress type is randomly selected. Then, one staircase process is randomly selected out of two staircases. Based on the previous response of a participant on that process, the next image is generated by applying stress of newly updated intensity and presented to the participant. Then the participant gives his feedback to that presented stressed image and requests the next one to present. This procedure lasts until all processes terminate. 10 participants attended in this experiment. The absolute threshold is selected at stress intensity getting 50% population vote on the psychometric function of each image.

Fig. 6 provides the results of the psychophysical experiment by describing the absolute threshold distributions as a box plot for each image of visual stress types. For images of R, SA, and SC stress types, the threshold values have narrow distributions, and threshold medians are also close to each other. For the other stress types, the threshold distributions are various, because the nature of how the applied stress changes symmetry pattern depends on the content of the image as well. However, for the majority of images, the perception thresholds are still close: For each of BH, BW, DW, DH, LH, NH stress types, the medians of thresholds of 4 images out of all five images have close values. For each of NW and LW, the thresholds of 3 images out of all five images are close to each other.

Experiment 2: Absolute Threshold for Each Visual Stress Type

The primary goal of this experiment is to detect the human perception threshold for each visual stress type. Unlike Experiment 1, which runs separate staircases for each image of visual stress type, in experiment 2, we create two interleaved staircases for each visual stress type. The maximum number of trials (presentations) for the progress is 50. For each visual stress type, at each presentation, stressed image
is randomly selected and presented with stress intensity that is calculated based on the participant’s response to the previously presented stressed image. In other words, all images of a particular visual stress type share the same staircase process. Over 40 participants attended this experiment in total, and for each visual stress type, thresholds of over 25 participants are collected. Fig. 7 illustrates a box plot that describes the distribution of threshold values for each visual stress type.

The absolute thresholds for BH, BW, and NH have broader distributions (high variance) than others. The average of the median thresholds in experiment 1 is consistent with those in experiment 2. We use the threshold values of experiment 2 for building our evaluation threshold. See figure 8 to observe sample stressed images that can be perceived as symmetry by humans.

IV. EVALUATION FRAMEWORK
Conventional evaluation of the reflection symmetry detection methods does not provide necessary insight into their behaviors on various visual stresses. However, the evaluation technique, which can demonstrate the limitations and advantageous aspects of detection methods, is beneficial and helpful in their further improvement. This kind of evaluation can also help to point out the applications where they can play

FIGURE 6: Threshold Distribution for each image of visual stress types.

FIGURE 7: Distributions of participants’ absolute thresholds for various visual stress types.
best. In this section, we propose the evaluation technique that contains the properties mentioned above. The first proposing evaluation technique evaluates the performance of symmetry detection methods over increasing visual stress intensity. This technique shows how the evaluating method reacts to the visual stress types. The second evaluation technique that we propose unleashes the detection limits (the best possible performance) of the evaluating methods on particular visual stress type.

Before moving to the detailed descriptions of these techniques, let’s introduce necessary notations and performance measures. Denote the symmetry decision rule $SDR$ as a function of $\tau$, and call $\tau$ a symmetry decision threshold. The symmetry decision threshold $\tau$ indicates the decision boundary between symmetry and non-symmetry ones of stressed symmetry images. In other words, given a reflection symmetry image, $\tau$ indicates the maximum intensity of visual stress in the image at which the stressed image still keeps its symmetry property. An image with stress intensity above $\tau$ has no symmetric pattern and is considered as non-symmetric. We can write the formulation of $SDR$ function as follows:

$$SDR(I, \tau) = \begin{cases} 1, & \text{if } d_I \leq \tau \\ 0, & \text{if } d_I > \tau \end{cases}$$

where $I$ is the stressed image and $d_I$ is the intensity of stress applied to the image $I$. Using the rule $SDR$ one can separate stressed images of specific visual stress type into two categories: $S$ - symmetry set and $NS$ - non-symmetry set: $S = \{I | SDR(I, \tau) = 1\}$ and $NS = \overline{S} = \{I | SDR(I, \tau) = 0\}$.

Let’s denote the set $D$ of stressed images that are detected as symmetry by the symmetry detection method. Note that evaluation ignores all detections that are not labeled. Then, given set $D$, the formulation of performance parameters, the true positive, false positive, true negative and false negative, look follows:

$$TP = S \cap D$$
$$FP = NS \cap D$$
$$TN = S \cap \overline{D}$$
$$FN = SN \cap \overline{D}$$

**Evaluation Technique 1.** The first evaluation technique is about demonstrating the behavior of symmetry detection methods over increasing stress intensity. For this evaluation, the symmetry decision threshold is fixed at human symmetry perception threshold. In order to show the performance trend, performance score is calculated at each visual stress intensity. F1-score as a performance measure might include redundant calculations. Because, if we denote the $L$ as a set of images that have the same intensity of a specific visual stress type, then $L$ is a subset of either set $S$ or set $NS$. If images of $L$ belong to $S$, then the set $D$ of images detected as symmetry is equal to true positives $TP$. Calculating precision $Pr$ is redundant (always 1) and makes the F1-score represent only recall $Rr$ (also called true positive rate $TPR$). Similarly, if images of set $L$ belong to $NS$, then precision $Pr$ is still redundant as there is no true positive $TP$ detection, and the recall is also undefined because of empty $S$. So, in this case, true negative rate, $TNR = \frac{|TN|}{|TN|+|FP|}$, is used as a performance measure. So, in second evaluation, we use true detection rate $TDR$ as performance measure:

$$TDR = \begin{cases} TPR, & \text{if } L \subset S \\ TNR, & \text{if } L \subset NS \end{cases}$$

**Evaluation Technique 2.** The second evaluation technique focuses on analyzing the performance of the computational symmetry detection method under various symmetry decision thresholds ($SDR$) of specific visual stress type. It also determines the best possible performance and its corresponding decision threshold. In order to measure the performance of symmetry detection method on a specific stress type, F1-score is used:

$$F1 = \frac{2 \cdot Pr \cdot Rc}{Pr + Rc}$$
V. EVALUATION RESULTS

In this section, using the proposed evaluation techniques, we evaluate three state-of-the-art methods: Reflection Symmetry Detection via Appearance of Structure Descriptor [61], Wavelet-based Reflection Symmetry Detection via Textural and Color Histograms [62], and Detecting Symmetry and Symmetric Constellations of Features [68]. These are top three well-performed methods in single 2D symmetry detection based on symmetry competition in ICCV 2017 workshop [26]. In order to detect symmetry patterns, Atadajanov and Lee [61] introduce appearance of structure features, which uses edges and contours in neighborhood. The method proposed by Elawady et al. [62] extracts edge/corner features using Log-Gabor filter, and describes them by their color and texture information in order to find reflected features. The method proposed by Loy and Eklundh [68] uses SIFT features for symmetric pattern detection.

For correct symmetry detection, we use the same rule provided in [28]. Two threshold values, \( t_1 \) and \( t_2 \), are used to determine correct detection. Successful detection happens, if the angle \( \theta \) between detected symmetry axis \( s \) and provided ground truth \( g \) is smaller than first threshold \( t_1 \) and the distance \( \delta \) to ground truth symmetry axis from the center of detected symmetry axis is smaller than second threshold \( t_2 \). In this evaluation \( t_1 = 3 \) deg and \( t_2 = 0.025 \) min (height, width).

First, we evaluate the behavior of the computational symmetry detection methods by using the evaluation technique 1. Second, we determine the performance trends and the best possible performance values by using the evaluation technique 2.

A. BEHAVIOR EVALUATION OVER INCREASING SYMMETRY STRESS INTENSITY

Before moving to the analysis, let’s define some terms (nouns and adjectives) that we use to describe the behavior of the methods in the context of proposed evaluation framework. Following are the necessary behavior properties which are vital to describe the performance and its behaviors.

- **Performance score** - a numerical value that indicates the performance of the detection method (e.g., true detection rate TDR).
- **Regularity** - a term describing common (expected or regular) behavior of the performance trend; non-increasing trend of performance measure is considered regular. Naturally, the expected trend of the performance measure is non-increasing as the intensity of stress increases. The opposite trend behavior is described by adjective, irregular.
- **Stability** - a term describing the frequency of the change in performance trend. A trend having fewer fluctuations (no fluctuation in an ideal case) is considered stable. So, the opposite trend performance behavior is fluctuating or unstable.

Our human annotated dataset provides the human perception threshold for various visual stresses. Based on these thresholds, images can be separated into two categories, "positive" (symmetry) and "negative" (non-symmetry). Fig. 9 illustrates the behavior of detection methods defined by their true detection rate (TDR), which is equal to either true positive rate, TPR, (Sensitivity) or true negative rate, TNR, (Specificity). The trend of TDR along visual stress intensity is presented. Unlike human perception, which usually has monotonic behavior on these visual stress intensities, most of the computational methods get the non-monotonic and fluctuating behavior. Behavior-wise, the method proposed by Elawady et al. [62] seems more stable than the others; it has a small variance of TDR along stress intensity. However, the method proposed by Loy and Eklundh [68] achieves overall the highest trend despite getting TDR level fluctuated. The method provided by Atadajanov and Lee [61] has very diverse TDR values throughout the stress intensity space. Below, we analyze the results for each visual stress type separately.

**Blurring.** Based on the results of human perception experiment, blurring creates a strong perceptual visual stress to reflection symmetry. Performance of the method provided by Atadajanov and Lee [61] drops almost to the half its initial performance with only small amounts of blurring applied. Detection method provided by them uses edge-based and contour-based features that describe the appearance of the image structure. Blurring the image fades the edges out and create new edges. Therefore, when the blurring applied to one half of the symmetry pattern, the strength of structure on that half decreases dramatically and achieves not similar descriptions. Contrarily, the features used by Loy and Eklundh [68] are SIFT. In consequence, this method gets the highest TDR scores. For images with the whole content blurred, the method proposed in [62] achieves the most stable TDR score in symmetry detection. In this evaluation, the methods proposed in [68] and [62] achieve stable and relatively high TPR till reaching the human symmetry perception threshold. For negative detection part, methods proposed in [61] achieves the highest and stable TDR.

**Brightness.** Changing the brightness of the image decreases the strength of both structure and appearance. However, changing brightness does not create new edges or change the location of features. Hence, the method given in [61] achieves much stable TDR values, unlike the blurring stress. However, the method still performs worse than the method introduced in [68] in the positive (symmetry) detection job. The method given in [62] achieves the worst TDR on BH stress type. In the non-symmetry detection job, the methods proposed in [61] and [68] perform competitively equal and get the highest TDR. For DW and BW stress type, the method provided in [62] achieves the best performance near the human perception threshold values. It makes the method provided by Elawady et al. [62] the most robust against brightness changing stress on the whole image (BW...
and DW. For BH stress type, the methods proposed by Loy and Eklundh [68], and Atadjanov and Lee [61] yield competitive performances.

**Noise.** Stressing the symmetry image by adding white noise is one of the most challenging visual stress types for all methods. Performance of methods proposed in [68] and [61] drop to zero before reaching half strength of noise which is applied to half of the symmetry pattern (NH). However, the method introduced in [62] achieves a non-zero performance in much larger intensity range before stress reaches the threshold value. It also has more stable behavior with having its performance decreasing gradually. For NW visual stress type, the method proposed in [68] achieves the highest performance. For DW, the method proposed by Elawady et al. [62] achieves the worst performance against NW stress type. For NW stress type, the method proposed by Elawady et al. [62] achieves competitive performances.

**Size.** The performance of all three methods drop to zero before reaching the threshold. Generally, the method proposed by Loy and Eklundh [68] has a high trend of performance values. Performance trend of the method has only two fluctuations which make the method the second best in being regular and stable. The performance trend of the method proposed in [61] has the most unstable and unpredictable form. The method yielded the worse performance on average for the stress type. The method of Elawady et al. [62] achieves a more regular and stable trend of all three trends.

**Skewness.** Human symmetry perception is robust to viewpoint change. Therefore, the psychophysical experiment achieved a high skew threshold. For all methods, the performance is dropped to zero before the stress intensity reaches...
the threshold in SC stress type. The method provided in [68] has the most irregular and fluctuating trend but the highest performance score. The method proposed by Atadjanov and Lee [61] has a more regular trend but the lowest performance score. The method introduced in [62] also has a regular trend.

B. PERFORMANCE EVALUATION ON VARIOUS SYMMETRY DECISION BOUNDARY

We evaluated the performance of the computational methods for each visual stress type by varying the symmetry decision threshold over stress intensity. Fig. 10 illustrates the performance trend and peak performance for each of three methods over the symmetry decision threshold axis. Based on the trend, the method proposed by Loy and Eklundh [68] achieves high performance over a larger portion of the abscissa axis. It also has higher peak value on all visual stress types except DH. However, for BH, NH, R and SA visual stress types method provided by [61] achieves higher performance on the lower values of the symmetry decision threshold. For only perfect (more perfect and less stressed) symmetry detection application, the method proposed in [61] achieves the highest performance. Table 1 provides exact performance scores with the symmetry decision threshold at the no-stress threshold, full-stress threshold and human perception threshold. The method provided in [68] achieves the best overall performance with the symmetry decision threshold at the full-stress threshold and the human perception threshold. However, the method provided by Atadjanov and Lee [61] performs the best on the no-stress threshold. The method proposed by Elawady et al. [62] outperforms other two methods in DW and NH stress types at high values of symmetry decision threshold. For stress type SC, all three methods get relatively close performance scores over all symmetry decision thresholds. Considering the whole symmetry decision space and all visual stress types, the method provided in [68] is the best, while the method provided in [62] is the second best in performance measure among three symmetry detection methods.

VI. CONCLUSION

In this work, we proposed a novel evaluation framework for computational symmetry detection methods based on human symmetry perception. The proposed framework evaluates the robustness and behavior of computational reflection symmetry detection methods on various visual stresses. Initially, we determined human symmetry perception limits on 11 visual stress types. For that, we conducted a psychophysical experiment. We, psychophysically, showed that the thresholds for each visual stress types are consistent with individual thresholds of images for those visual stress types. We introduced modifications to the up-down staircase method and developed a web-based system to conduct the psychophysical experiment. Based on human perception limits, we built a human annotated dataset for all 11 stress types with various stress intensities and introduced necessary performance measures for the proposed evaluation framework. We evaluated three state-of-the-art computational reflection symmetry detection methods using the proposed framework. The proposed evaluation framework showed how the evaluating methods are robust to various visual stress and behavior of evaluating methods as a function of stress intensity. In our view, the proposed evaluation framework provides more increased insight into the weak and strong aspects of the evaluating methods than the traditional evaluations could do.

REFERENCES

performance scores with the symmetry decision threshold achieves the highest performance. Table 1 provides exact symmetry detection application, the method proposed in [61] for only perfect (more perfect and less stressed) visual stress types method provided by [61] achieves higher abscissa axis. It also has higher peak value on all visual [68] achieves high performance over a larger portion of the methods over the symmetry decision threshold axis. Based performance trend and peak performance for each of three symmetry decision threshold values.

**FIGURE 10:** F1-score of computational reflection symmetry detection methods for each visual stress type with various visual stress types and all visual stress types, the method considering the whole symmetry decision space and all visual stress types, the method for stress type SC, all three two methods in DW and NH stress types at high values of F1-score.

**TABLE 1:** F1-score of computational reflection symmetry detection methods for three symmetry decision thresholds.

<table>
<thead>
<tr>
<th>Visual Stress Types</th>
<th>No-stress Threshold</th>
<th>Full-stress Threshold</th>
<th>Human Perception Threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brightening Half</td>
<td>0.40</td>
<td>0.18</td>
<td>0.09</td>
</tr>
<tr>
<td>Blur Whole</td>
<td>0.17</td>
<td>0.29</td>
<td>0.15</td>
</tr>
<tr>
<td>Darkening Half</td>
<td>0.09</td>
<td>0.14</td>
<td>0.143</td>
</tr>
<tr>
<td>Darkening Whole</td>
<td>0.13</td>
<td>0.14</td>
<td>0.09</td>
</tr>
<tr>
<td>Brightening Half</td>
<td>0.08</td>
<td>0.21</td>
<td>0.00</td>
</tr>
<tr>
<td>Brightening Whole</td>
<td>0.07</td>
<td>0.13</td>
<td>0.08</td>
</tr>
<tr>
<td>Noise Half</td>
<td>0.29</td>
<td>0.182</td>
<td>0.173</td>
</tr>
<tr>
<td>Noise Whole</td>
<td>0.00</td>
<td>0.22</td>
<td>0.00</td>
</tr>
<tr>
<td>Size</td>
<td>0.26</td>
<td>0.11</td>
<td>0.14</td>
</tr>
<tr>
<td>Skew Along Symmetry Axis</td>
<td>0.44</td>
<td>0.14</td>
<td>0.16</td>
</tr>
<tr>
<td>Skew Across Symmetry Axis</td>
<td>0.11</td>
<td>0.14</td>
<td>0.11</td>
</tr>
<tr>
<td>All Stimuli</td>
<td>0.159</td>
<td>0.157</td>
<td>0.12</td>
</tr>
</tbody>
</table>


IBRAGIM R. ATADJANOV received the B.S. degree in Applied Mathematics and Computer Science from Urgench State University, Urgench, Uzbekistan, in 2008, and the M.S. degree in Information Security from Tashkent University of Information Technologies, Tashkent, Uzbekistan, 2011. He is currently pursuing the Ph.D. degree in Computer Science and Engineering at Kyung Hee University, the Republic of Korea.

His research interests include feature detection and description, pattern recognition, symmetry-based computer vision, multilayer compressive light field displays and deep learning.

SEUNGYU LEE received the B.S. and M.S. degrees in electrical engineering from the Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 1997 and 1999, respectively, and the Ph.D. degree in computer science & engineering from The Pennsylvania State University in 2009.

He has been a Research Engineer with the Technical Research Institute, Korea Broadcasting System, Seoul, Korea, where he has been involved in the research on high-definition image processing, MPEG4 advanced video coding, and the standardization of terrestrial-digital mobile broadcasting. He has also been a Principal Research Scientist with the Advanced Media Laboratory, Samsung Advanced Institute of Technology, Yongin, Korea. He is currently an Associate Professor with Kyung Hee University, Seoul. His research interests include time-of-flight depth camera, color/depth image processing, symmetry-based computer vision, and 3-D modeling and reconstruction.

***