Gaze-Wasserstein: A Quantitative Screening Approach to Autism Spectrum Disorders

Kun Woo Cho, Feng Lin, Chen Song, Xiaowei Xu, Michelle Hartley-McAndrew,
Kathy Ralabate Doody, and Wenyao Xu

***Abstract*— Early detection of children with autism spectrum disorder (ASD) has been of great interest to researchers due to an increase in the rate of autism incidence around the world. However, a diagnosis of ASD is still challenging to receive in a timely manner for the large-scale population because the current diagnostic practice requires considerable cost and time, and do not provide quantitative feedback. In this paper, we explore a new ASD screening method, namely *Gaze-Wasserstein*, that is non-invasive, fast, and widely accessible. Based on the gaze tracking and analysis, Gaze-Wasserstein is able to provide objective gaze pattern-based measurements for home-based ASD screening, and can eventually be deployed on any mobile technologies with a front camera. To test the performance of Gaze-Wasserstein, we conducted a pilot study with 32 child participants where 16 children have ASD and 16 children are typically developing. Evaluation results demonstrate the effectiveness and time-efficiency of our proposed method in the ASD screening, which indicate that our Gaze-Wasserstein is a promising autism screening approach in the clinical practice.**

I. Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental condition that is defined by concerns in three major domains: social interaction, communication, and behavior [1]. The impairment in social interaction causes an abnormality in many nonverbal behaviors related to eye-to-eye gaze, facial expression, and body gestures [2]. Also, restricted repetitive patterns of behavior are often manifested by the persistent and intensive preoccupation with parts over whole.

As autism is reported to occur in all racial, ethnic, and socioeconomic groups, its prevalence is about 1-2 per 1000 people worldwide. According to the Centers for Disease Control and Prevention (CDC) [3], about 1 in 68 children in the United States has been diagnosed with ASD in 2016 while a government survey of parents suggests that 1 in 45 children has been identified with ASD. This signifies that 2 percent of children in the U.S. are living with autism, which is notably higher than the official estimates reported by CDC [4]. This gap between two reports implies the significant limitation on the current ASD diagnostic system.

K. Cho, F. Lin, C. Song, X. Xu, W. Xu are with the Department of Computer Science and Engineering, University at Buffalo (SUNY), Buffalo, New York 14260 USA (corresponding author: Wenyao Xu; phone: 716-645-4748; fax: 716-645-3464; e-mail: {kunwooch, flin28, csong5, xiaoweix, wenyaoxu}@buffalo.edu)

M. Hartley-McAndrew is with Children's Guide Foundation Autism Spectrum Disorder Center, Buffalo, NY 14222 USA (e-mail: hartley4@buffalo.edu)

K. Doody is with the Department of Exceptional Education, Buffalo State University (SUNY), Buffalo, New York 14222 USA (e-mail: doodykr@buffalostate.edu)

To date, the gold standard in diagnosing autism is called the Autism Diagnostic Observation Schedule (ADOS) [5]. However, the ADOS requires implementation by specialized clinical settings and trained professionals, which makes it costly and inefficient, preventing a timely ASD diagnosis for a large population. Furthermore, the current practices in measuring ASD behavioral markers are still subjective as the accuracy highly depends on the expertise and experience of physicians. Due to these limitations, the average age of ASD diagnosis in the U.S. is approximately 5 years old [6] while most of the autistic children begin to exhibit specific behavioral markers as early as the first to second year [7].

In fact, the early detection of autism is necessary for the pre-emptive educational planning and treatment, provision for family supports and education, and delivery of appropriate medical care [8]. Particularly if a diagnosis can be made earlier, appropriate therapy can encourage child's malleable brain to reroute around faulty neural pathways [9]. Therefore, an objective and evidence-based screening approach for ASD is urgently needed.

Recently, several studies [10], [11] point out that human attention (e.g., gaze) is a promising marker of the early diagnosis of ASD. In particular, the gaze behavior is closely related to human attention as Dakin $et\ al.$ [12] have found the abnormality in visual function of individuals with ASD on the visual perception stimulus.

Also, Song $et\ al.$ [13] explore the abnormality in the dynamic gaze pattern of children with ASD when they process a given social situation. The result indicates that this abnormality is caused by their lack of ability to understand the relationship depicted in the social scene. Based on these works, we further explore the gaze pattern in both social scene and non-social scene for the ASD screening.

In this paper, we present a novel gaze-based ASD screening method, *Gaze-Wasserstein*, by incorporating a modified 1st Wasserstein distance for the dissimilarity measure. As the Wasserstein distance is stochastic and associated with the gradient flow, 1st Wasserstein distance has advantages of shorter computation time, higher accuracy, and higher robustness to noises from irrelevant stimuli. Due to these factors, our approach is also suitable for implementation of ASD diagnosis system in mobile technologies. To the best of our knowledge, there is no study in the literature to explore advanced distance metrics for ASD screening systems. In our system, k nearest neighbors classification (kNN), f-score (F1) accuracy, equal error rate (EER), and receiver operating characteristics curve (ROC) are further employed

978-1-5090-3090-3/16/$31.00 ©2016 IEEE

to comprehensively evaluate the system performance. With 32 participants (16 children with ASD and 16 typically developing children), the average of recall and precision rate achieves 94.17% and 93.75%, and f-measure accuracy achieves 93.96% for social scene stimulus.

In summary, there are three contributions in this paper:

- An analysis of new ASD screening method, named Gaze-Wasserstein, that employs 1st Wasserstein distance as a dissimilarity measure for discrete gaze distribution.
- A comparison of the system performance on two types of visual stimulus (social scene and non-social scene).
- A validation of feasibility of our proposed method by performing a pilot study.

II. Related work

Gaze behavioral markers of ASD have already been explored by many previous studies. First, many studies [12], [14] have proved a restrictive and unique gaze pattern of the children with ASD. For example, Senju *et al.* [15] compared direct gaze and averted gaze stimulus among typically developing children and children with ASD. The result showed that children with ASD were better with detecting averted gaze than detecting direct gaze, which was unlike normal children.

Second, several literature have demonstrated how the impairment in social interaction may affect the gaze pattern of children with ASD. For instance, Rutherford *et al.* [16] investigated how individuals with ASD focus on the facial feature when perceiving emotional expressions. When choosing an image that resembles pre-defined emotion in reality, individuals with ASD were more likely to select the most exaggerated facial expressions, which the participants without ASD thought grotesque and unnatural. Also, Pelphrey *et al.* [17] found that people with ASD spent longer time on viewing non-feature areas of the faces while spending less time on examining core features related to emotional expressions. In this study, participants with ASD showed a deficit in recognizing emotion, primarily due to their inability in recognition of fear.

Third, many researchers also have examined the abnormalities in other aspects during the emotion and face recognition task. Pierce *et al.* [18] discovered that although subjects with ASD are able to somewhat perform the face perception task, none of the region in the brain that supports the face processing were found to be significantly active. Similarly, Mammarella *et al.* [19] examined the performance of children with ASD and typical development (TD) control in visuospatial working memory (VSWM) tasks under different complexity. High semantic and low semantic visual spatial stimuli were both presented. The result showed that TD group were advantaged with the high semantic stimuli and group of children with ASD had a detail-focused processing style and were unable to utilize long-term memory semantics to construct global representations of the array. Additionally, some studies [20] [21] demonstrated that autism group showed different response in GSR upon emotional visual stimuli, while some others [22] hold a second opinion in some conditions.

In sum, previous researches have convincingly proved the early differences of gaze behavior between people with and without ASD, particularly when parsing the behavior induced by social symptomatology. This, in turn, provides multiple ways for the detection of ASD. While the existing biofeedback based screening methods [23], [24], [25] require highly controlled laboratory environment, we present more efficient and feasible ASD screening approach for the implementation in mobile technologies.

III. Materials

A. Participants

Our study was approved by the Institutional Review Board of Women & Children Hospital, SUNY, University at Buffalo and Buffalo State [IRB: 595026-3]. Collection of data is obtained from 32 participants ranging in age from 2 to 10 year. Of the participants, 19 were male (59%) and 13 were female (31%). 16 children (9 of the males and 7 of the females) have been previously diagnosed with ASD according to the Autism Diagnostic Observation Schedule (ADOS) and criteria provided by Diagnostic Statistic Manual (DSM). Other 16 participants without ASD (10 of the males and 6 of the females) are classified as typically developing (TD) children who have not been diagnosed with any neurodevelopmental disorders. Every participants were recruited through an existing research program and parental consents were obligatory and were obtained at the time of the study. All participants, including children and their families, have received a comprehensive description of the experiment and its requirements.

B. Hardware Unit

Collection of data is acquired with a Tobii EyeX Controller [26], which tracks the gaze pattern in response to the visualization of stimulus. Particularly, the Tobii EyeX Controller utilizes near-infrared light to measure the movement of eyes and detects the x and y axis of the gaze point at the frequency of 120 Hz. Its operating range is 18" $\times$ 40" with an eye to application latency of 15 $ms \pm 5$ ms. The screen size can be up to 27" and its weight is 0.2 lb with a head-box size of 16" $\times$ 12". Fig. 1 shows how Tobii EyeX Controller works on the visual stimulus.

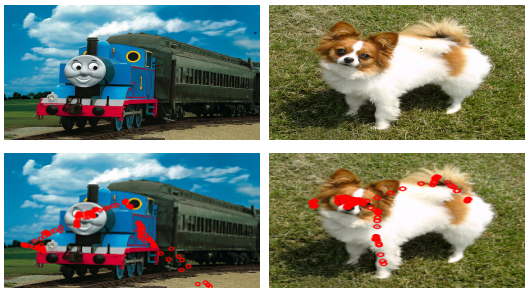


Fig. 1. Top row shows the examples of the visual stimulus and bottom row shows the examples of the gaze pattern detected by Tobii EyeX.

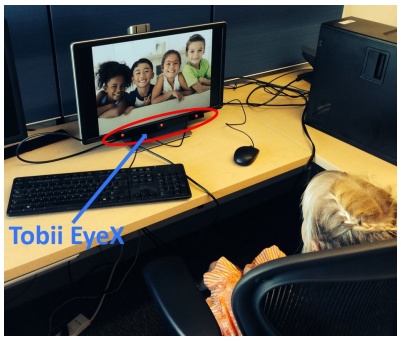


Fig. 2. A child is responding to the visual stimulus while the Tobii EyeX controller detects the gaze behavior.

C. Procedures

Prior to the experiment, children and families were provided with photographs of the research location, as well as pictures of the researchers and graduate assistants, in order to get children familiarized with the environment setting. Also, reinforcers, such as snack bars, drinks, and stickers, were offered to participants to encourage compliant behavior and awarded at the completion of the data collection session. During the experiment, participants were guided into a small room and seated at a table with computer equipment. The task was organized in the series of *eight* visual stimuli in total. Each pre-designed images was displayed to the participants for five seconds and shifting to next image took two seconds. Thus, total experimental process took approximately 54 seconds. For selection of the visual stimulus, we have chosen the images that have a visually clean background in order to prevent any unintentional distraction from irrelevant stimuli. Fig. 2 shows one of the children participating the experiment.

IV. Our Gaze-Wasserstein Framework

The Gaze-Wasserstein screening system is shown in Fig. 3. This figure illustrates two main components of the Gaze-Wasserstein: local access and remote access. Data of discrete gaze distribution are acquired through a device that the child is locally accessing. Although data collection in this experiment is acquired by desktop, data can be collected through mobile technologies, such as smart-phones, tablets, and wearable eye tracker (see Section VI in detail). Then the data are being remotely accessed and analyzed by physicians. Details of data analysis and screening are discussed in this section.

A. 1st Wasserstein Distance

1st Wasserstein distance (WD) [27] is a very natural way to compare the probability density functions of two variable P and Q, where P is derived from Q by small, nonuniform perturbation. Compared to deterministic distance metrics, 1st Wasserstein distance has shorter computation time [28] and it finds dissimilarity more accurately when the ground distance is perceptually meaningful [29]. Moreover, it is insensitive to oscillations and, therefore, allows our model to be intrinsically robust to noises from accidental eye movement.

Let two distribution P and Q represent as follow: first gaze point distribution $P = \{(x_1, p_1) \cdots (x_m, p_m)\}$, $1 \leq i \leq m$, and second gaze point distribution $Q = \{(y_1, q_1) \cdots (y_n, q_n)\}$, $1 \leq j \leq n$; and $D = [d_{ij}]$ is a ground distance matrix where d_{ij} is the ground distance between clusters x_i and y_j. Flow $F = [f_{ij}]$ is the solution of [27], [29]:

$$W_p(P, Q, F) = min(\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij} d_{ij}) \tag{1}$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij} d_{ij} = \sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij} \left\| x_i - y_j \right\|^p = E_F \left\| X - Y \right\|^p \tag{2}$$

with subject to

$$f_{ij} \geq 0; 1 \leq i \leq m, 1 \leq j \leq n \tag{3}$$

$$\sum_{j=1}^{n} f_{ij} = p_i; 1 \leq i \leq m \tag{4}$$

$$\sum_{i=1}^{m} f_{ij} = q_j; 1 \leq j \leq n \tag{5}$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij} = \sum_{i=1}^{m} p_i = \sum_{j=1}^{n} q_j = 1 \tag{6}$$

First constraint forces the supplies to move from P to Q but not from Q to P [30]. Second constraint shows that the amount of supplies that can be sent by the clusters in P is equal to its weights. Then, third constraint shows that the clusters in Q from receiving supplies is equal to its capacity. The last constraint forcefully moves the maximum amount of supplies, which is also referred as a total flow. Once the transportation of supplies is done, 1st Wasserstein Distance is defined as [27]

$$W_p(P, Q) = \min_{F} \left\{ E_F \left\| X - Y \right\|^p \right\}^{1/p} \tag{7}$$

where p is value greater or equal to 1. $\|.\|$ indicates the L_p vector norm. Minimum of the expectation is taken over joint probability distributions F when the marginal distribution of X is P and the marginal of Y is Q.

However, gaze distributions of children with ASD often contain very few gaze points compared to that of typically developing (TD) children. It primarily because the children with ASD often show a restricted gaze behavior. When these very few points are within the boundary of the gaze distribution of TD that is being compared, original 1st Wasserstein algorithm perceives the distribution of ASD gaze pattern as a part of the gaze distribution of TD control and, thereby, gives a partial matching and misclassifies the subject with ASD to TD subject. In order solve this issue, our 1st

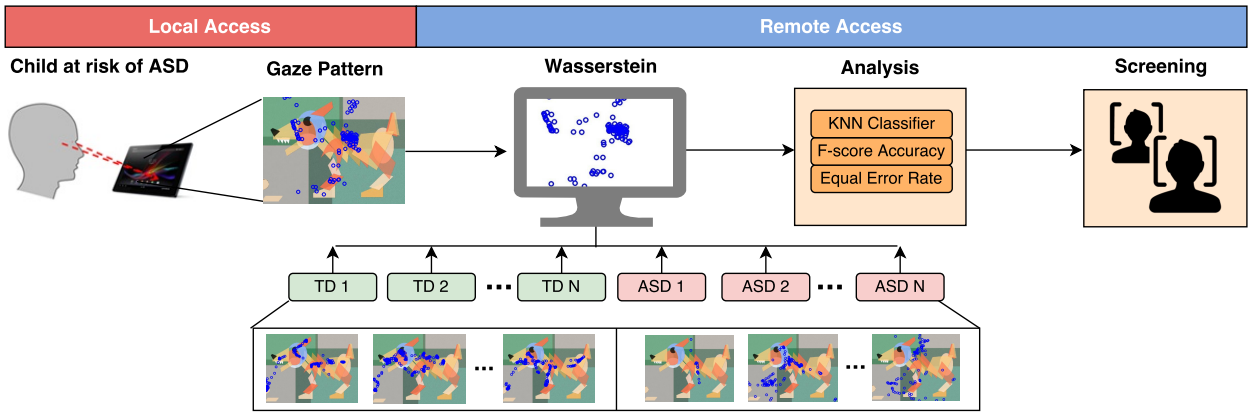


Fig. 3. The Gaze-Wasserstein Autism Screening Framework.

Wasserstein distance adds a penalty to the difference in the number of points as follow [31]:

$$\widehat{W}_{p_\alpha}(P,Q) = \min_{f_{ij}} \sum_{i,j} f_{ij} d_{ij} + \left| \sum_i P_i - \sum_j Q_j \right| \times \alpha \cdot \max_{i,j} \{d_{ij}\} \quad (8)$$

This algorithm shares the same constraints with the original 1st Wasserstein distance. If the masses are not equal, $\widehat{W}_p$ makes masses on both side to become equal by adding one supplier or demander. The ground distance between these added elements to other demanders or suppliers is set to be α times the maximum ground distance [31]. For Gaze-Wasserstein, we employed the α value of 0.5.

B. Classification with k Nearest Neighbors

With k-nearest neighbors (kNN) classifier, an object is classified by a majority (a positive integer k) vote of its neighbor. [32]. Data set $S = \{(x_i, y_i)\}$ where $x_i \in \mathbb{R}^p$ and $y_i \in \{1, \cdots, j\}$. Testing data are $x \in \mathbb{R}^p$. We must determine the label of x, which can be represented as y. For our study, $x_i \in \{W_{p_{0,0}} \cdots W_{p_{n,n}}\}$; $y_i \in \{TD_1 \cdots TD_{16},\ ASD_1 \cdots ASD_{16}\}$; $a = W_{p_{testing}}$. In this study, we setup the value of k as 3 (see Section V-C in detail).

V. Performance Evaluation

A. Evaluation Description

There are *eight* visual stimuli given to each of 32 child participants. The images that are illustrating more than one human figure are classified as a social scene (Fig. 4(a)) and the images depicting only few non-human figures are considered as a non-social scene (Fig. 4(b)). There are four social scene stimuli (SS) and four non-social scene stimuli (NSS). ASD-SS, TD-SS, ASD-NSS, and TD-NSS respectively represent ASD group experimented on the social scene stimuli, TD group with the social scene stimuli, ASD group with the non-social scene stimuli, and TD group with the non-social scene stimuli. In addition, ASD-total is composed of ASD-SS and ASD-NSS data, and TD-total is composed of TD-SS and TD-NSS data.

Leave-one-out-cross-validation (LOOCV) is further applied to each eight trials. For the test with SS stimulus, one of the gaze distribution on SS, whether it is a gaze distribution of ASD subject (ASD-SS) or gaze distribution of TD subject (TD-SS), is selected as a target data while the rest ASD-SS and TD-SS data are considered as a training set. Then kNN classification is done. This process repeats for each 32 subjects. For the test with NSS stimulus, one of NSS data, whether it is ASD-NSS or TD-NSS, is selected as a target data while the rest ASD-NSS and TD-NSS are regarded as a training set. Then kNN classification is done. Again, this process repeats for 32 times.

(a) Social scene stimuli (SS). (b) Non-social scene stimuli (NSS).

Fig. 4. Eight visual stimuli where four depict social scenes in (a) and other four depict non-social scenes in (b).

B. Evaluation Results

1) Accuracy: In order to have a comprehensive analysis of the system accuracy, we first employ the f-measure accuracy or balanced f-score (F_1), which is known as a harmonic mean of precision p and recall r. p is the number of true positive (TP) divided by the number of positive calls (TP+FP) while r is the number of true positive (TP) divided by the number of condition positives (TP+FN) where FP is false positive and FN is false negative. Simply, F1 is defined as follow:

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} = \frac{2TP}{2TP + FP + FN} \quad (9)$$

Overall performance (ASD-total, TD-total) for all eight trials (four SS and four NSS) is summarized in Table I. Its average precision and recall are 91.41% and 92.08% with the f-measure accuracy of 91.74%. This implies that our system can correctly classify ASD and TD child subjects with the accuracy of **91.74%** when utilizing both SS and NSS.

Table II shows a performance of the system on social scene alone. It has the average precision and recall value of 93.75% and 94.17% with the f-measure accuracy of **93.96%**. On the other hand, the performance of the system on non-social scene is shown in Table III. Its average precision and recall value are 89.06% and 89.99% along with the f-measure accuracy of **89.52%**. In all aspects, the performance for SS stimulus is better than that of the overall performance and that of the performance for NSS stimulus alone. It is important to point out that f-score accuracy of SS is higher than f-score accuracy of overall by 2.22% and f-score accuracy of NSS by 4.44%. This 4.44% suggests that utilizing social scene for Gaze-Wasserstein is recommended over using non-social scene.

TABLE I

OVERALL PERFORMANCE TABLE

Total Scene	Recall (%)	Precision (%)	EER (%)
ASD	87.50±8.84	95.14±3.03	6.28±4.83
TD	95.31±2.89	89.02±7.14	5.35±3.03
Average	91.41±3.64	92.08±3.14	5.82±3.92

TABLE II

PERFORMANCE TABLE FOR SOCIAL SCENE

Social Scene	Recall (%)	Precision (%)	EER (%)
ASD	92.19±7.86	95.41±3.07	3.40±2.67
TD	95.31±3.13	92.93±6.52	3.19±2.06
Average	93.75±2.55	94.17±2.08	3.30±1.28

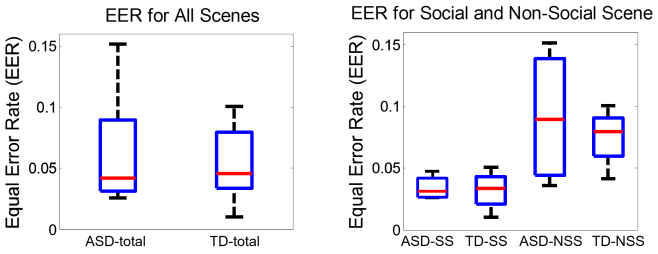
TABLE III

PERFORMANCE TABLE FOR NON-SOCIAL SCENE

Non-Social Scene	Recall (%)	Precision (%)	EER (%)
ASD	82.81±7.86	94.87±3.45	9.15±5.60
TD	95.31±3.13	85.10±5.97	7.51±2.48
Average	89.06±3.13	89.99±2.66	8.33±4.10

2) Equal Error Rate: Equal error rate (EER) is a measure to evaluate the system performance. By definition, it is a rate where the acceptance error, known as TPR, is equal to the rejection error, known as FNR. The accuracy of the system is high when EER value is low. Unlike the accuracy metric, EER indicates the sensitivity of false positive and false negative on the ASD screening. Box-plot of EER values for ASD and TD is shown in Fig. 5.

Figure. 5(a) illustrates the EER value for overall performance in both SS and NSS. The error bar represents the standard deviation of EER. EER value of ASD-total is 6.28% and TD-total is 5.35% with the standard deviation of 4.83% and 3.03%, correspondingly. In total, the average of EER value for both ASD and TD is 5.82%.



(a) EER for both social and non-social scene.

(b) Separate EER for social and non-social scene. SS implies social scene and NSS implies non-social scene.

Fig. 5. EER of ASD and TD. The standard deviation is represented by the error bars.

Figure. 5(b) shows the EER value for SS and NSS, separately. EER value of ASD-SS and TD-SS are 3.40% and 3.19% with the standard deviation of 2.67% and 2.06% while that of ASD-NSS and TD-NSS are 9.15% and 7.51% with the standard deviation of 5.60% and 2.48%, respectively. As shown in Fig. 5(b), EER value of SS is always less than the value of NSS for all cases. Thus, EER value for SS is small enough to conclude that the Gaze-Wasserstein is highly robust and has good screening sensitivity when utilizing the social scene.

C. K-value Selection on the Classification Accuracy

For this part, we investigate the impact of k-value for kNN classification on the performance of our system. As mentioned above, our k-value for the Gaze-Wasserstein is 3. In this experiment, we examine six k-values (1, 3, 5, 7, 9, and 11) to explore an appropriate k-value for our system.

All k-value tests are experimented on both SS and NSS (average of all eight scenes where four images are SS and four images are NSS) and the results are illustrated in Fig. 6. As shown in Fig. 6, the classification accuracy increases from $k = 1$ to $k = 3$ and gradually decreases from $k = 3$ to $k = 11$. This trend indicates that $k = 3$ provides the best system accuracy. In addition, the standard deviation of f-measure accuracy for $k = 3$ (3.37) is the smallest among the standard deviations for all k-values. This implies that our system is highly stable when utilizing $k = 3$ for kNN classification. Thus, the appropriate k-value for Gaze-Wasserstein is 3.

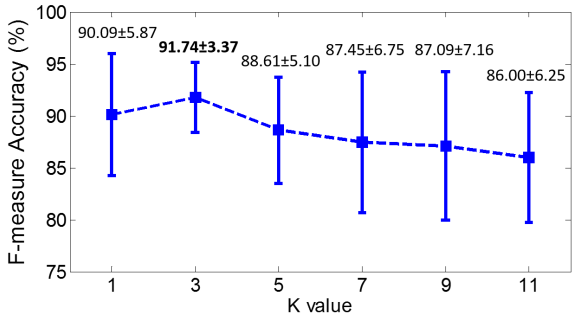


Fig. 6. Impact of k-value selection on the classification accuracy. The standard deviation is represented by the error bars.

D. Optimization of Screen Time Efficiency

As previously mentioned, each of pre-designed visual stimuli is displayed for $five$ seconds and shifting to next visual stimulus takes two seconds. The pilot study is composed of four SS stimuli and four NSS stimuli. Therefore, the total screen time took approximately 54 seconds. It is very important to optimize the total time of the ASD screening because child participants may have low compliance and limited attentions in the study.

We investigate the time efficiency by reducing the number of visual stimuli in the screening process. SS and NSS stimulus are examined separately and the results are depicted in Fig. 7. Initially, we start from one stimulus which requires 5 seconds. Then, we add the number of stimulus by one until there are four stimuli that requires 26 seconds of the screening process. As shown in the figure, the screening accuracy (F-measure) increases as the screen time scales up for both SS and NSS setups. Also, f-measure accuracy at 5-second screen time for SS can still reach over 91%. This result demonstrates that our proposed screen time is efficient to adopt in the real ASD screening practice.

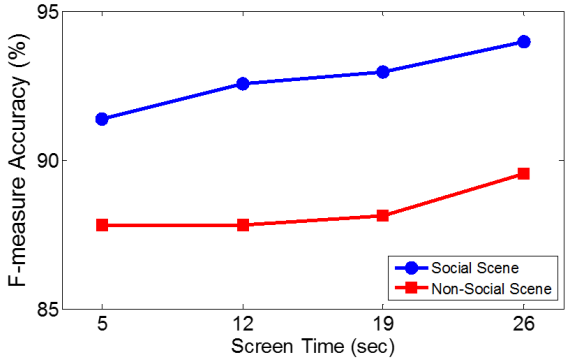


Fig. 7. Impact of screen time on the performance of Gaze-Wasserstein.

E. Comparison with Other Distance Metrics

1) Earth Mover's Distance: Table IV depicts the recall and precision of the ASD screening system using Earth Mover's Distance (EMD) [30]. EMD is known to be exactly same as 1st Wasserstein when two distributions have the same total mass [29], [33], [27]. However, our discrete gaze pattern distributions do not have the same total mass. Also, EMD allows the partial matching of gaze distributions while our modified 1st Wasserstein distance does not allow it.

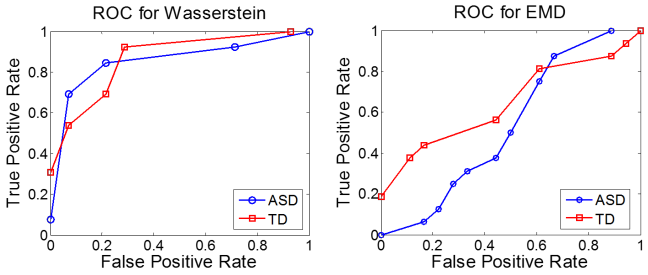
TABLE IV

OVERALL PERFORMANCE TABLE FOR EMD

Total Scene	Recall (%)	Precision (%)
ASD	0.00 ± 0.00	0.00 ± 0.00
TD	94.44 ± 6.41	48.55 ± 1.70
Average	47.22 ± 3.21	24.28 ± 0.85

For performance using EMD, we have averaged out the result of four scenes (two social scene and two non-social scene) with exactly same participant as before (32 children). Using EMD, every children with ASD are classified as TD subject for all four trials and, therefore, ASD classification had 0% accuracy. As a result, total recall, precision, and f-score value are respectively 47.22%, 24.28%, and 32.05%. However, there is only 0.87% difference between EMD's recall value of TD (94.44%) and the recall value of TD using 1st Wasserstein distance (95.31%). This implies that partial matching characteristic of EMD does not affect TD classification as much as ASD classification.

Receiver operating characteristic curve (ROC) is further applied to effectively visualize the performance difference between EMD and 1st Wasserstein distance. By definition, ROC is calculated by a true positive rate (sensitivity) against false positive rate (fall-out) at the various threshold settings. Test performance is considered to be more accurate as the curve follows the left-top portion of the space. Both Fig. 8(a) and Fig. 8(b) are experimented on the SS stimulus. As shown in Fig. 8(a), the performance of ASD and TD for 1st Wasserstein are very similar to each other and their curves cross each other over different setups. For EMD, the performance of TD, as illustrated in the Fig. 8(b), is better than the performance of ASD because the partial matching characteristic made ASD subjects to be classified as TD subjects. Observing the general performance of two figures, using 1st Wasserstein distance gives higher accuracy for both of ASD and TD than using the EMD. This result corresponds to the previous results on the recall, precision, and f-measure accuracy.

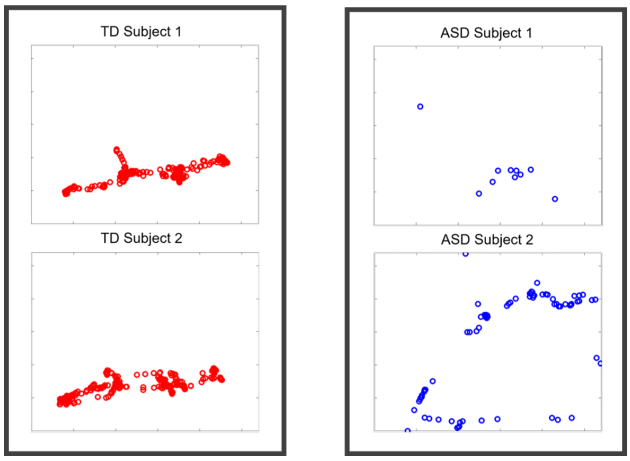


(a) ROC of ASD and TD using Wasserstein distance that does not allow partial matching.

(b) ROC of ASD and TD using Earth Mover's Distance that does allow partial matching.

Fig. 8. ROC of ASD and TD for EMD and Wasserstein.

Partial matching characteristic in EMD is ineffective for the screening of ASD because many gaze distributions of children with ASD have a very small number of gaze fixation points. Since the ASD subjects have restricted gaze patterns and tend to show a persistent preoccupation with parts over whole, their range of view can be fixed to only few parts of the object, producing less number of points compared to TD controls in the same period of time (top of Fig. 9(b)). However, there are cases where the impairment in social interaction plays a larger role. In this case, ASD subjects lose the focus and the gaze points often become very randomly and widely distributed as shown in the bottom figure of Fig. 9(b). As a result, the randomness causes a larger dissimilarity value when computed with another wide and random distribution of ASD subject.



(a) Gaze distribution of TD subjects.

(b) Gaze distribution of ASD subjects.

Fig. 9. Gaze distribution of four subjects. All four gaze patterns are based on the same SS image.

2) Hausdorff Distance: The Hausdorff Distance (HD) [34] is known as a minimum bound of Euclidean distance (ED). Let two finite point sets represent as follow: first set $A = \{a_1, \cdots, a_m\}$ and second set $B = \{b_1, \cdots, b_n\}$. Then the Hausdorff distance is described as [35]

$$H(A, B) = max(h(A, B), h(B, A)) \tag{10}$$

$$h(A, B) = \underset{a \in A}{max} \underset{b \in B}{min} \|a - b\| \tag{11}$$

Here, $h(A, B)$ is known as the directed Hausdorff distance from set A and B with L_2 norm $\|.\|$ on the points of A and B [35]. By definition, the Hausdorff distance is the longest distance of all the distances from a point in one set to the closest point in the other set. Table V depicts the recall and precision of ASD screening system using Hausdorff distance. In addition, f-score of Hausdorff distance is 75.16%. Its precision, recall, and f-score are less than that of 1st Wasserstein distance by 17.19%, 15.95%, and 16.56%. Thus, overall performance with Hausdorff distance is relatively lower than the overall performance with 1st Wasserstein distance.

TABLE V
OVERALL PERFORMANCE TABLE FOR HAUSDORFF DISTANCE

Total Scene	Recall (%)	Precision (%)
ASD	60.94 ± 5.98	83.06 ± 6.19
TD	87.50 ± 5.10	69.19 ± 3.96
Average	74.22 ± 4.69	76.13 ± 4.86

Summary: Figure 10 demonstrates the advantages of using 1st Wasserstein distance over EMD and Hausdorff distance. When compared with EMD, 1st Wasserstein distance is highly accurate for the screening of children with ASD due to the absence of the partial matching characteristic. In comparison with Hausdorff distance and other deterministic metrics, stochastic 1st Wasserstein distance provides a better performance due to its robustness to outlier gaze points.

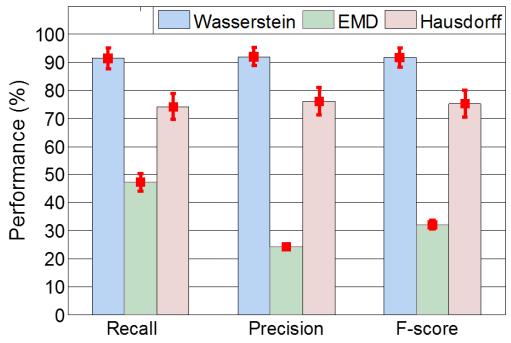


Fig. 10. Performance comparison among Wasserstein, EMD, and Hausdorff distance. Error bar represents the standard deviation.

VI. Discussion

Our Gaze-Wasserstein screening system could potentially be work on the portable hardware, such as smart-phones and tablets, without the need for additional eye tracking devices like Tobii EyeX Controller [26]. In order to perform Gaze-Wasserstein algorithm, only two technologies are needed: a monitor to display visual stimuli and hardware containing the eye-tracking system. In fact, eye-tracking technology using mobile devices have already been developed in the past. For instance, Song *et al.* presented a novel eye-movement based authentication system for smart-phones. Using facial info pre-processing and gaze-angle calculation, a front camera of the phone extracts gaze pattern that reflects both physiological and behavioral aspects in nature [36]. Similarly, Krafka *et al.* introduced an eye tracking solution targeting mobile devices. As the handsets camera captures user's face, a software, named iTracker, accounts the position and direction of the head and eyes to determine the position of gaze fixation point [37]. Thus, current mobile technologies satisfies all aspects for Gaze-Wasserstein with several benefits over other platforms. One advantage is that advanced multitasking system in mobile technologies allows the device to display visual stimuli while operating front camera. Also, a fixed position of the front camera relative to the screen decreases the number of unknown parameter and, thereby, provides the high-accuracy calibration-free tracking [37]. Other benefits include accessibility and portability.

However, our technique can also be applied to wearable eye trackers [38]. In this case, the visual stimulus is replaced from images to live view. Thus, benefit of using wearable eye tracker will be that it could offer an accurate data of what a child is looking at, wirelessly and in real time, and, therefore, provide immediate response in the actual social situations.

VII. Conclusion

In this paper, we presented a gaze pattern-based ASD screening approach, called Gaze-Wasserstein. Unlike previously published works, Gaze-Wasserstein enhances the global gaze pattern matching by employing the modified 1st Wasserstein distance. kNN classification and leave-one-out-validation are applied to validate the effectiveness of our

approach. The evaluation results showed that our approach achieves the average recall and precision of 94.17% and 93.75%, and f-measure accuracy of 93.96% for social scene stimulus. This study not only demonstrates the feasibility of our ASD screening approach in the clinical practice, but also opens the way to implement early ASD diagnosis system in the mobile technologies.

VIII. Acknowledge

The authors would like to thank our shepherd Dr. Benny Lo and anonymous reviewers for their insightful comments on this paper.

References

[1] Lorna Wing and Judith Gould. Severe impairments of social interaction and associated abnormalities in children: Epidemiology and classification. *Journal of Autism and Developmental Disorders*, 9:11–29, 1979.

[2] American Psychiatric Association et al. Diagnostic and statistical manual of mental disorders american psychiatric association. *Washington, DC*, 210, 1994.

[3] Centers for Disease Control and Prevention. ASD Data and Statistics. `http://www.cdc.gov/ncbddd/autism/data.html`. accessed by May 9, 2016.

[4] B Zablotsky, LI Black, MJ Maenner, LA Schieve, and SJ Blumberg. Estimated prevalence of autism and other developmental disabilities following questionnaire changes in the 2014 national health interview survey. *National health statistics reports*, 2015.

[5] Katherine Gotham, Andrew Pickles, and Catherine Lord. Standardizing ados scores for a measure of severity in autism spectrum disorders. *Journal of autism and developmental disorders*, 39(5):693–705, 2009.

[6] Paul T Shattuck, Maureen Durkin, Matthew Maenner, Craig Newschaffer, David S Mandell, Lisa Wiggins, Li-Ching Lee, Catherine Rice, Ellen Giarelli, Russell Kirby, et al. Timing of identification among children with an autism spectrum disorder: findings from a population-based surveillance study. *Journal of the American Academy of Child & Adolescent Psychiatry*, 48(5):474–483, 2009.

[7] Lonnie Zwaigenbaum, Susan Bryson, Tracey Rogers, Wendy Roberts, Jessica Brian, and Peter Szatmari. Behavioral manifestations of autism in the first year of life. *International journal of developmental neuroscience*, 23(2):143–152, 2005.

[8] Geraldine Dawson. Early behavioral intervention, brain plasticity, and the prevention of autism spectrum disorder. *Development and psychopathology*, 20(03):775–803, 2008.

[9] Deborah Fein, Marianne Barton, Inge-Marie Eigsti, Elizabeth Kelley, Letitia Naigles, Robert T Schultz, Michael Stevens, Molly Helt, Alyssa Orinstein, Michael Rosenthal, et al. Optimal outcome in individuals with a history of autism. *Journal of Child Psychology and Psychiatry*, 54(2):195–205, 2013.

[10] Warren Jones and Ami Klin. Attention to eyes is present but in decline in 2-6-month-old infants later diagnosed with autism. *Nature*, 2013.

[11] Lauri Nummenmaa, Andrew D Engell, Elisabeth von dem Hagen, Richard NA Henson, and Andrew J Calder. Autism spectrum traits predict the neural response to eye gaze in typical individuals. *Neuroimage*, 59(4):3356–3363, 2012.

[12] Steven Dakin and Uta Frith. Vagaries of visual perception in autism. *Neuron*, 48:497–507, November 2005.

[13] Chen Song, Aosen Wang, Kathy Doody, Michelle Hartley-McAndrew, Jana Mertz, Feng Lin, and Wenyao Xu. Analyzing dynamic components of social scene parsing strategy in autism spectrum disorder. In *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, pages 132–135. IEEE, 2016.

[14] Rutherford MD and Towns AM. Scan path differences and similarities during emotion perception in those with and without autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 38(7):1371–81, August 2008.

[15] Atsushi Senju. Eye contact does not facilitate detection in children with autism. *Cognition*, 89(1):B43–B51, August 2003.

[16] MD Rutherford and Daniel N McIntosh. Rules versus prototype matching: Strategies of perception of emotional facial expressions in the autism spectrum. *Journal of autism and developmental disorders*, 37(2):187–196, 2007.

[17] Kevin A Pelphrey, Noah J Sasson, J Steven Reznick, Gregory Paul, Barbara D Goldman, and Joseph Piven. Visual scanning of faces in autism. *Journal of autism and developmental disorders*, 32(4):249–261, 2002.

[18] Karen Pierce, R-A Müller, J Ambrose, Greg Allen, and Eric Courchesne. Face processing occurs outside the fusiformface area'in autism: evidence from functional mri. *Brain*, 124(10):2059–2073, 2001.

[19] Irene C. Mammarella. Visuospatial working memory in children with autism: The effect of a semantic global organization. *Research in Developmental Disabilities*, 35(6):1349–1356, June 2014.

[20] RJR Blair. Psychophysiological responsiveness to the distress ofothers in children with autism. *Personality and Individual Differences*, 26(3):477–485, 1999.

[21] Anneli Kylliäinen and Jari K Hietanen. Skin conductance responses to another person's gaze in children with autism. *Journal of autism and developmental disorders*, 36(4):517–525, 2006.

[22] D Ben Shalom, SH Mostofsky, RL Hazlett, MC Goldberg, RJ Landa, Y Faran, DR McLeod, and R Hoehn-Saric. Normal physiological emotions but differences in expression of conscious feelings in children with high-functioning autism. *Journal of autism and developmental disorders*, 36(3):395–400, 2006.

[23] Esubalew Bekele, Zhi Zheng, Amy Swanson, Julie Crittendon, Zachary Warren, and Niladri Sarkar. Understanding how adolescents with autism respond to facial expressions in virtual reality environments. *Visualization and Computer Graphics, IEEE Transactions on*, 19(4):711–720, 2013.

[24] Kim M Dalton, Brendon M Nacewicz, Tom Johnstone, Hillary S Schaefer, Morton Ann Gernsbacher, HH Goldsmith, Andrew L Alexander, and Richard J Davidson. Gaze fixation and the neural circuitry of face processing in autism. *Nature neuroscience*, 8(4):519–526, 2005.

[25] Karen Pierce, David Conant, Roxana Hazin, Richard Stoner, and Jamie Desmond. Preference for geometric patterns early in life as a risk factor for autism. *Archives of general psychiatry*, 68(1):101–109, 2011.

[26] Tobii AB. Tobii eyex controller. http://www.tobii.com/xperience/products/Specification. Accessed by 05/17/2016.

[27] Ludger Rüschendorf. The wasserstein distance and approximation theorems. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 70(1):117–129, 1985.

[28] Kangyu Ni, Xavier Bresson, Tony Chan, and Selim Esedoglu. Local histogram based segmentation using the wasserstein distance. *International journal of computer vision*, 84(1):97–111, 2009.

[29] Elizaveta Levina and Peter Bickel. The earth mover's distance is the mallows distance: some insights from statistics. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 2, pages 251–256. IEEE, 2001.

[30] Yossi Rubner, Carlo Tomasi, and Leonidas J Guibas. The earth mover's distance as a metric for image retrieval. *International journal of computer vision*, 40(2):99–121, 2000.

[31] Ofir Pele and Michael Werman. A linear time histogram metric for improved sift matching. In *Computer Vision–ECCV 2008*, pages 495–508. Springer, 2008.

[32] Leif E Peterson. K-nearest neighbor. *Scholarpedia*, 4(2):1883, 2009.

[33] Ofir Pele and Michael Werman. Fast and robust earth mover's distances. In *Computer vision, 2009 IEEE 12th international conference on*, pages 460–467. IEEE, 2009.

[34] Daniel P Huttenlocher, Gregory A Klanderman, and William J Rucklidge. Comparing images using the hausdorff distance. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 15(9):850–863, 1993.

[35] Oliver Jesorsky, Klaus J Kirchberg, and Robert W Frischholz. Robust face detection using the hausdorff distance. In *Audio-and video-based biometric person authentication*, pages 90–95. Springer, 2001.

[36] Chen Song, Aosen Wang, Kui Ren, and Wenyao Xu. Eyeveri: A secure and usable approach for smartphone user authentication. *IEEE International Conference on Computer Communication*, 2016.

[37] Kyle Krafka, Aditya Khosla, Petr Kellnhofer, Harini Kannan, Suchendra Bhandarkar, Wojciech Matusik, and Antonio Torralba. Eye tracking for everyone. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2176–2184, 2016.

[38] Tobii AB. Tobii pro glasses 2. http://www.tobiipro.com/product-listing/tobii-pro-glasses-2/. Accessed by 05/17/2016.