Automated Autism Detection based on Characterizing Observable Patterns from Photos

Alice Z. Guo

***Abstract*—Autism spectrum disorder (ASD) is a developmental disorder that affects the communication and behavior. People with ASD show atypical attentions to social stimuli and gaze at human faces and complex scenes in an unusual way, and their facial expressions are often atypical as well. This work investigates the feasibility of developing an automated method to analyze the visual cues of autism using the photos taken by people with ASD, comparing to photos taken by people without ASD, in different scenarios. It was inspired by a recent study based on manual inspection of the photos. The key challenge is what and how to characterize the photos taken by people with ASD, to facilitate an automated separation from normal people. Several features are proposed to characterize the observable behaviors for ASD with experimental validations. This is the first work to perform an automatic analysis of the photos taken by people with ASD, achieving a prediction accuracy of 85.8%.**

***Index Terms*—Autism spectrum disorder (ASD), automated ASD detection, ASD from photos, visual cues, feature extraction, attention and expression mined from photos**

I. Introduction

Autism spectrum disorder (ASD) was first described by Kanner in 1943 as a disorder characterized by an inability to have a substantial impact on the long-term relationships with other people, delayed speech, language abnormalities, an obsessive desire for sameness, and onset in early infancy [1]–[3]. It is a life-long developmental disorder characterized by qualitative impairments in social and communication behavior and a restricted range of activities and interests [4], [5]. People with ASD show atypical attention to social stimuli and gaze at faces and complex scenes in unusual ways [6]–[8]. ASD is estimated to affect 1 in 150 persons, thereby it is no longer considered as a rare disorder [9]. Many primary care pediatricians (PCPs) care for people with ASDs, especially children. It is critical that PCPs can recognize the signs of ASDs as soon as possible. Rapid advances in the fields of cognitive and affective developmental neuroscience, developmental psychopathology, and applied behavior analysis have contributed new methods for early ASD detection. Currently, ASD diagnoses are made mainly based on the clinician judgment with specific criteria as a guide. Thus there is a practical demand to develop automated methods for ASD detection.

Several studies have established [10] that facial expressions of children with autism are often perceived as atypical, awkward or less engaging by typical adult observers. Based on these, the computational analysis of children with ASD can be, for instance, based on using a high functioning autism (HFA) device for detailed facial expression capture and analysis [10], a virtual reality (VR) augmented eye tracker for gaze capture [11] or with a multi-sensory instrument for gaze [12], [13], and using computer vision techniques for analysis [14], [15]. In addition to visual analysis, there are also approaches using the online blogs to detect the autism people via the extraction of affective information from their posts and comments in blogs [16], which can be useful for analysis of adults. In contrast to previous works, a new computational approach is explored to analyze the photos taken by the autism people, motivated by a recent finding [17] that people with autism take photos differently from the control or normal people.

Alice Z. Guo is with Morgantown High School, Morgantown, WV 26505, USA. Email: aliceguo168@gmail.com.

Wang et al. [17] studied people with ASD by asking them to take photos in several scenarios. This was a new viewpoint to study ASD. They found that people with ASD may take photos differently from normal people. However, their approach utilized a manual check and examination of the taken photos, which may constrain its application to a small scale data set, because of the long processing time and labor cost. Therefore, it will be useful if an automated method can be developed to analyze the photos taken by people with ASD. The automated technique could be applied broadly to promptly detect ASD patients in a large population.

To the best of knowledge, this is the first time an automated method is explored and developed for ASD people detection based on characterizing the observable behaviors or patterns from the photos they took.

In the remainder of this paper, Section II presents what properties can be considered for photos taken by people with ASD, and how to extract effective features to characterize these properties from the photos. Section III describes the employed learning method for classification between ASD and normal people. Section IV provides the data set and experimental results with some discussions and insights on the proposed approach. Finally, conclusions are drawn in Section V.

II. Feature Extraction

To develop an automated method for autism detection from photos, the key issues are what properties should be characterized from the photos, and how to represent said characteristics. Based on [17] and a careful examination of the data, four properties are proposed to be considered: (1) repetitive photos; (2) blurred photos; (3) tilted photos; and (4) unusual portrait photos, which include incomplete faces, occluded bodies, and photos taken without the subjects being asked to pay attention to the camera.

In the following, the specific methods will be presented to characterize these properties separately, aiming at transforming the described observable properties into computational algorithms. An additional benefit is that these features can be used

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2020.3035088, IEEE Transactions on Affective Computing

to explain the symptoms for ASD diagnosis, not just for ASD classification.

A. Repetitive photos

One property of people with ASD is that they may take photos repeatedly. Those photos are similar, but not the same, e.g., with movements or changes of the camera poses/locations, while capturing the scene repeatedly. Characterizing these photos is different from duplicate image recognition [18] in computer vision, where the images are about the same with some minor image manipulations (not camera changes). The Gist feature [19] is chosen to extract a kind of global representation of the images, while ignoring local details. It is computed by convolving an image with filters at different scales and orientations. The high and low frequency repetitive gradient directions of an image can be computed. The scores for filtering at various orientations and scales are computed as Gist features.

Repetitive photos were captured by the subjects with ASD in each scenario, but not all of the photos are repetitive. To measure the percentage of repetitive photos captured by each subject, pairwise comparisons are considered. The similarity scores (or cosine distances) are calculated between the image pairs. To determine if one image is repetitive with others, it is compared to all remaining photos, and the maximum value of the similarity scores is computed. This is done for each image captured by the subject, and a histogram of similarity scores with a fixed number of bins for quantization (e.g., 13 in the experiments). The histogram feature can handle the various numbers of images taken by different subjects. Fig. 1 shows two histograms, one is calculated from the photos taken by a subject with ASD, another by a subject without ASD. The histograms of pair-wised similarities are used to represent the repetitive photos. The main process of computing the histogram feature is given in Algorithm 1.

Algorithm 1 Pseudo code to compute repetitive histograms.

Input: K images captured by a subject.
Output: Computed N histogram bins for the subject.

for each image **do**
    Extract Gist feature
end for
Calculate M pairs of images for the K images
for $i = 1$; $i \leq$M; $i++$ **do**
    Compute the cosine distance (cos_dis_i) of Gist features
end for
for $j = 1$; $j \leq$K; $j++$ **do**
    Select the maximum similarity (cos_dis_j) for j-th image
end for
Compute a histogram feature $F = [f_1, ..., f_N]$ from the K similarities

B. Blurred photos

Another property is that people with ASD may often take blurred photos. To characterize these photos, the Tenengrad autofocus function is computed for each image. This algorithm convolves an image G with the Sobel operators and sums the square of all the magnitudes greater than a threshold [20]:

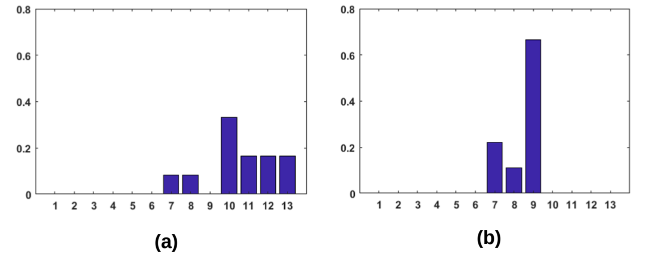

Fig. 1. Examples of histograms measuring the repetitive photos: (a) from the photos by a subject with ASD, and (b) by a subject without ASD.

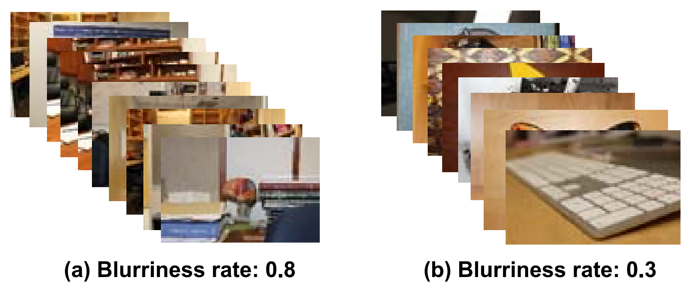

Fig. 2. Estimated blurriness rates for the photos from two subjects: (a) from a subject with ASD, and (b) from a subject without ASD.

$$F_{ten} = \sum_{i,j} |g(i,j) \otimes S|^2 + \sum_{i,j} |g(i,j) \otimes S'|^2, \quad (1)$$

where S and S' are the Sobel's kernel and its corresponding transpose, respectively:

$$S = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}. \quad (2)$$

With the blurriness measure, a threshold is set (e.g., 2.7 in the experiments) to determine if the image is blurred or not. Then, a blurriness rate is computed for the photos captured by a subject, which is the proportion of blurred images to the total number of images. Note that not all photos taken by people with ASD are blurred, that is why the rate of blurriness is computed. Fig. 2 shows two samples with different blurriness rates, taken by a subject with ASD and another subject without ASD, respectively.

C. Tilted photos

The third property involves tilted photos, as an example shown in Fig. 3 (a). To find the tilted photos, a Radon transform [21] based method is developed to characterize the angles of straight lines. The method has three main steps. A detailed description of the method is given below.

Given an image $I(x,y)$, the Radon transform is defined as:

$$R_\theta(\rho) = \int_{x_i}^{x_j} \int_{y_m}^{y_n} I(x,y)\delta(\rho - xcos\theta - ysin\theta)dxdy, \quad (3)$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2020.3035088, IEEE Transactions on Affective Computing

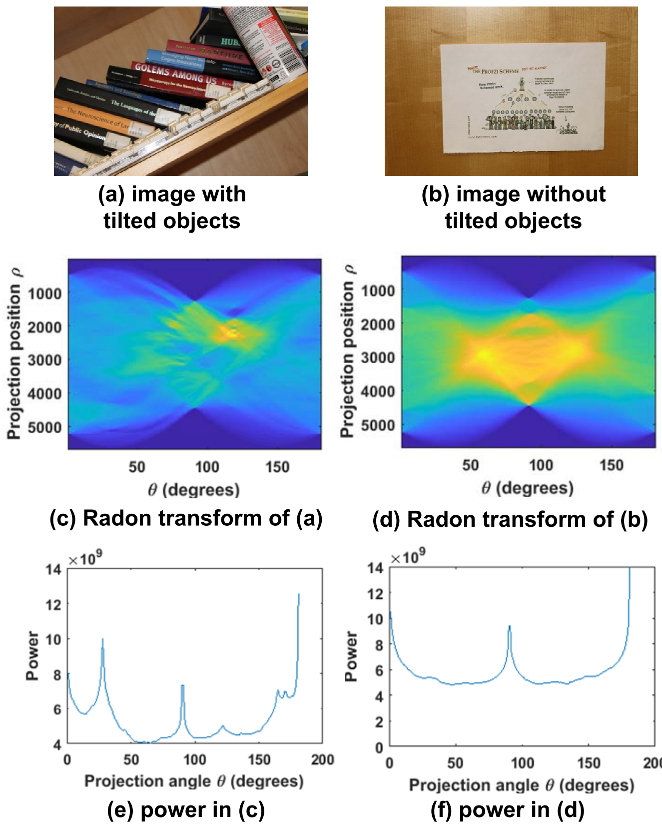

Fig. 3. Characterizing tilted photos vs. normal photos.

where θ is the angle of a projection, ρ is the offset of a projection line from the origin of the (x, y) plane, x_i, x_j, y_m, y_n define four boundary points of the (x, y) plane, and $\delta(\cdot)$ is the Dirac delta function.

Then, Fourier transform [22] is applied to each projection parameterized by the angle. Fourier transform for a projection of an image at angle θ can be found as:

$$r_\mu = \int_{-\infty}^{\infty} R_\theta exp(-j2\pi\mu\rho)d\rho, \quad (4)$$

were μ represents a value along the y axis.

The decision rule is based on the power distribution. The total power of all projections between 80 and 100 degrees (i.e., around 90 degrees with an allowable deviation $\Delta = \pm 10$ degrees) is compared to the total power of all projections above 100 and below 80 degrees, and a decision is made for the presence or absence of tilted lines. Determined as tilted if

$$\sum_{\theta=20^\circ}^{80^\circ} |r_\theta|^2 + \sum_{\theta=100^\circ}^{160^\circ} |r_\theta|^2 > \gamma \sum_{\theta=80^\circ}^{100^\circ} |r_\theta|^2, \quad (5)$$

where γ is a decision threshold ($\gamma > 1$). The image is not tilted, otherwise. The power for below 20 or above 160 degrees is discarded, because of some possible boundary noise. Fig. 3 shows two images, tilted (left) and not tilted (right). Their power spectra shown in Fig. 3 (e) and (f) are distinctive to separate between them.

Finally, a tilt rate is computed for all images captured by each subject, which is the proportion of tilted images to the total number of images. Again, not all photos are tilted when taken by a person with ASD. The tilt rate is used to characterize the set of photos taken by a subject with ASD, and separate from the set of photos taken by a normal subject.

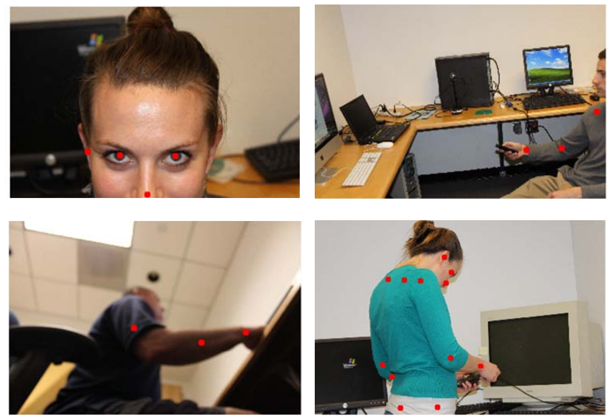

Fig. 4. Portrait photos with detected skeleton-joints marked as red points.

D. Unusual portrait photos

The last property is that people with ASD take unusual photos of humans, called unusual portrait photos. It includes incomplete human faces, occluded or partial bodies, and the photos taken without asking the subjects to pay attention to the camera. These unusual portrait photos may be caused by the atypical attention of people with ASD, and impairments in social and communication behavior. To characterize these photos, the human body skeleton-joints can be detected first, using a toolbox called OpenPose [23]. Fig. 4 shows some portrait photos with detected skeleton-joints. For a complete human body, there are about 20 joints that can be detected by the tool. The number of detected joints and the joint locations can be used to determine if a photo contains incomplete faces, profile faces, or bodies without heads:

(1) An incomplete face can be determined by computing how many facial landmarks are detected, and what landmarks are detected. If some major facial landmarks, e.g., eyes, nose, and mouth, cannot be found in the photo, the face is considered incomplete. (2) Profile faces can be determined based on the head pose angle. If the head pose is far from a frontal view, e.g., close to 90 degrees rotation, it is considered as a profile face picture. (3) For a detected human body, if the head region is not included within the photo, it is considered as a body without head.

For autism people, part of the photos that they take may have these unusual portrait properties. The ratios of these photos over the total number of photos they take should be computed for the characterization. Thus all three rates, i.e., incomplete face rate, profile face rate and the body without head rate, are computed for the images captured by each subject, as the extracted features.

III. Learning the Classifier

A feature vector can be obtained by concatenating the above computed features together, which includes repetitive histograms, blurriness rate, tilted rate, incomplete face rate, profile face rate and without head rate. To determine if a subject has ASD or not, the binary support vector machine (SVM) [24] can be used to learn a classification function.

The SVM is a supervised learning algorithm [25], widely used in many problems. The SVM can do nonlinear classification using kernels. The Gaussian radial basis function (RBF) kernel is one of the most popular kernels for nonlinear SVM. In this work, the SVM with RBF kernel is adopted for the automated autism detection.

The SVM classifier usually does not require a large number of training examples, which is a nice property for our ASD detection problem, since it is often small for the captured dataset in developing algorithms.

IV. Experiments

The data set will be introduced first, and then experiments are conducted on the dataset, which evaluates the effectiveness of the proposed autism detection methods. Finally, some discussions and insights are presented.

A. Dataset

The dataset [17] contains 1988 photos taken by 37 participants. Among them, 16 participants have ASD, and 21 do not (denoted as the control); The ASD group had a full-scale IQ (FSIQ) of 111.6±12.2 (mean±SD, from the Wechsler Abbreviated Scale of Intelligence-2), a mean age of 29.7±11.2 years. The Controls group had a comparable full scale IQ of 111.0±9.90, and a comparable mean age of 33.0±9.31 years. Fig. 5 shows the age distribution of the participants in the dataset, ranging from 20 to 55 years old. The controls were also matched on gender, race and education to the ASD group [17].

Each participant may contribute one or two samples (based on visit times), resulting in 50 samples in total, as shown in Table I. Each sample contains a different number of photos, varying from nine to 53. The consumer digital camera, Canon EOS REBEL T1i, was used for each participant, who was told to use the 'auto' mode of the camera. No special training on photography was given to any of these participants. Further, all participants were surveyed about their previous experience in photography, and there is no significant difference between the two groups in terms of professional training of photography. There are three different scenes [17] for each participant: (1) indoors and of people in a lab; (2) indoors in the lab without people; and (3) outdoors. The outdoors photos were taken in the daytime. In each scene, the participants took photos at will. Fig. 6 shows some samples from the three scenarios.

In this dataset, each sample was rated by three independent raters who are familiar with the clinical presentation of ASD and are reliable on the standard called ADOS-2 (Autism Diagnostic Observation Schedule, 2nd Edition, Module 4) [26].

In the experiments, the samples from the same individual can only be in either training or testing, but cannot overlap. A standard 4-fold cross validation is executed for the experiments.

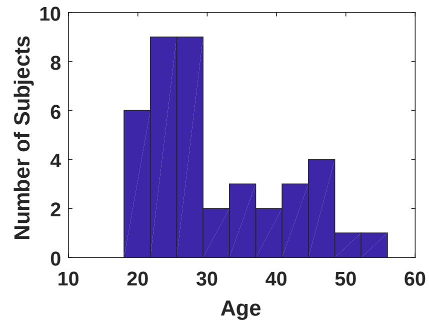

Fig. 5. The age distribution of the participants in the data set.

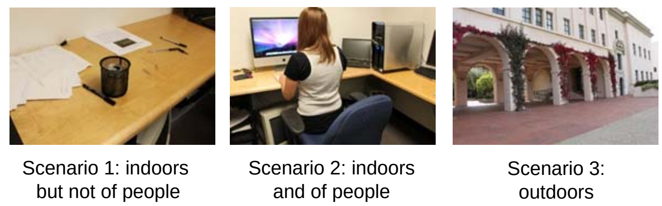

Fig. 6. Some photos to illustrate the three scenarios.

B. Performance Measure

The dataset introduced above is for the first time to be explored for a computational method, there is no previous work to tell how to measure the algorithm's performance. Take it as a pattern recognition problem, the recognition accuracy can be used. Further, it can be informative to check some details about the results, such as how many people with ASD are classified as normal, rather than just reporting the overall accuracy. As a result, three measures are computed:

(1) *False acceptance rate (FAR)*, the proportion of negative cases that are incorrectly classified as the positive, and the lower the better; (2) *True positive rate (TPR)*, the proportion of positive cases that are correctly classified, and the higher the better; and (3) *Accuracy*, the proportion of the total number of cases that are correctly recognized, and the higher the better.

In our case, the positive represents people with ASD. From the TPR, one can infer the false negative rate (FNR), with the relation FNR$=1-$TPR. Similarly, from the FAR, one can know the false positive rate (FPR), and then infer the true negative rate (TNR), by TNR$=1-$FAR. Given these, the confusion matrix and F1 Score can also be computed.

C. Experimental Results

The photos from two scenarios are investigated in this work: indoors but not of people, and indoors and of people. The outdoor scene [17] does not show clear properties to characterize in the examination, it is removed from experiments.

TABLE I
Details about the data set.

	ASD	Control
# of samples	21	29
# of images	1011	977

TABLE II
THE RECOGNITION RESULTS UNDER DIFFERENT SCENARIOS.

Scenario	Feature	FAR(%)	TPR(%)	Accu.(%)
Indoors, not of people	Repetitive, blurred, and tilted	34.4	95.0	78.0
Indoors, of people	Unusual portrait	44.7	92.1	70.2
All indoors scenarios	All features	24.1	100	85.8

TABLE III
RECOGNITION RESULTS BY SINGLE FEATURES: INDOORS, NOT OF PEOPLE.

Feature	FAR(%)	TPR(%)	Accuracy(%)
Repetitive	45.1	62.5	58.0
Blurriness	31.3	33.3	53.1
Tilted	51.8	38.3	46.2

TABLE IV
RECOGNITION RESULTS FOR UNUSUAL PORTRAITS: INDOORS, OF PEOPLE.

Sub-feature	FAR(%)	TPR(%)	Accuracy(%)
Incomplete face rate	7.15	19.2	61.9
Profile face rate	41.4	40.8	52.4
Without head rate	48.2	61.7	55.9

Table II depicts the detection results in different scenarios. The first column indicates the selected scenario, and the second denotes the features corresponding to each scenario. For the indoors but not of people scenario, the recognition accuracy is 78.0%. For the indoors and of people scenario, the recognition accuracy is 70.2%. When combining the two scenarios by concatenating the features together, the accuracy is raised to 85.8%. This accuracy is relatively high, showing the feasibility to develop a computational method to analyze the photos taken by people with ASD.

In addition to the recognition accuracies, which are usually measured for many pattern recognition problems, Table II also shows the FAR and TPR. In the specific dataset, the best result can have an TPR of 100%, while the FAR is 24.1%. As an error measure, the FAR is not very low, but is reasonably good. Certainly, there are rooms to improve the developed algorithms in future works, hopefully to reduce the errors further.

To get a detailed understanding of each specific feature, the performance of each single feature is measured for the two scenarios, as shown in Tables III and IV, respectively. For the scenario of indoors but not of people, the accuracies are different for the three different features. The repetitive feature can deliver an accuracy of 58.0%, higher than the other two features in Table III. For the scenario of indoors and of people, the portrait features include three sub-features: incomplete face rate, profile face rate, and the body without head rate, which perform differently. The feature of incomplete face rate shows an accuracy of 61.9%, higher than the other features in Table IV. Fusing all features together leads to a much better performance than using each single feature, i.e., achieving an accuracy of 85.8%, listed in the last row in Table II, showing that all developed features are needed. The corresponding confusion matrix is shown in Table V, and the F1 Score is 0.89, which is quite high.

TABLE V
CONFUSION MATRIX OF THE BEST ASD DETECTION RESULT.

		Prediction	
		ASD	Control
Actual	ASD	100%	0%
	Control	24.1%	75.9%

D. Discussion and Insights

The experimental results show the feasibility of developing a computational approach to analyze the photos taken by people with ASD, telling a fact that some special properties can be characterized computationally for autism people through their taken photos. As a result, how autism people "see" the other humans and the world can be recorded and analyzed. In some sense, this "see" is different from the use of eye tracking [11]–[13], which usually do not show blurred or occluded photos as in this study; The eye tracking may convey local saliency but not a global view as the taken photos in our case.

In previous studies, it has been shown that how autism people "say" about the world can be analyzed through their online blogs [16], while this study provides a new way to understand autism people from a different angle - photos. Thus the photos may be combined with the blogs or text information from autism people, in order to deliver a better understanding of people with ASD. This combination could be an interesting research topic. Our view is that exploring more sources of information can be useful to characterize the autism symptoms better, and eventually achieve an automated diagnosis of people with ASD accurately, objectively, rapidly, and economically.

In terms of the proposed feature extraction methods to characterize the photos taken by people with ASD, the starting point is to observe the photos carefully and find the special properties visually, and then develop appropriate methods accordingly. The experimental validation shows that the developed methods can work, demonstrating the feasibility to utilize their taken photos to distinguish people with ASD computationally, and indicating that some observable behaviors of Autism people can be "transformed" into computational methods, which encourages more careful observations of various behaviors from Autism people. Further, the computational approach developed in this work can explain why the autism spectrum disorder is detected, based on what symptoms are computed. As a result, the developed computational system can be made "believable," "explainable," and "understandable" to human experts for further or more careful examination and diagnosis.

Given the computational approach, however, one may ask: Why not using the deep learning methods?

Recently, the deep learning methods [27], [28] have shown great successes in many applications. However, there are several reasons why they are not used here: (1) typically the deep learning methods need a large number of training examples, which are difficult to provide in the current autism dataset. In future, if a large number of samples can be captured, there may be an opportunity to consider the deep learning methods; (2) In addition to dataset size, the deep learning methods are usually supervised learning, requiring the labels for each sample. However, as described in previous sections, there are no labels, such as repetitive photos or tilted photos, for each picture or set of pictures. On the other hand, if only the class information is provided, e.g., autism or control, it might be too difficult for deep learning methods to learn the relevant features without further annotations of specific behavior labels; (3) the deep learning methods usually work on kinds of "regular" datasets, while the autism dataset in our case is not that regular. For instance, there is neither a common basis to align these photos, like regular face or object images, nor common features for all photos taken by the same subject. That is why many of the features proposed here are computing the rates or ratios from a set of photos taken by the same person; (4) Even though the deep learning techniques could be used for autism spectrum disorder detection with a good accuracy, they usually cannot explain why the subject is detected as an Autism person or not. The lack of explanation problem may raise doubts to the patients, since it is difficult to persuade or convince them without explanations that they are diagnosed with the ASD. Therefore, even though the deep learning methods could be used with a good accuracy for ASD detection, the methods developed here are still useful to understand the key symptoms for ASD analysis, explanation, and diagnosis.

V. Conclusion

It has been investigated on the feasibility of developing an automated method to analyze the photos taken by people with autism spectrum disorder, in comparison to the photos taken by normal people. To the best of knowledge, this is the first work to explore a computational approach for ASD detection based on characterizing the observable behaviors from the photos taken by people with ASD. The well-known atypical attention and expression for people with ASD are depicted indirectly by the photos they take. The developed methods can characterize the photos for different properties in different scenarios, and explain why the autism spectrum disorder is detected through showing the discovered patterns that are observable to humans. The experimental validation on a real data set has shown that it is promising to develop an automated method for analyzing the photos taken by people with ASD. This study has opened a door for inspiring more advanced approaches for automated autism detection, with the potential for deployment on large scale applications.

Acknowledgment

The author thanks S. Wang for providing data for the study, and M. Jiang for help and discussion of the methods.

References

[1] J. A. Osterling, G. Dawson, and J. A. Munson, "Early recognition of 1-year-old infants with autism spectrum disorder versus mental retardation," *Development and psychopathology*, vol. 14, no. 2, pp. 239–251, 2002.

[2] S. E. Bryson, S. J. Rogers, and E. Fombonne, "Autism spectrum disorders: early detection, intervention, education, and psychopharmacological management," *Canadian J. of Psychiatry*, vol. 48, no. 8, pp. 506–516, 2003.

[3] C. J. Newschaffer, L. A. Croen, J. Daniels, E. Giarelli, J. K. Grether, S. E. Levy, D. S. Mandell, L. A. Miller, J. Pinto-Martin, J. Reaven *et al.*, "The epidemiology of autism spectrum disorders," *Annu. Rev. Public Health*, vol. 28, pp. 235–258, 2007.

[4] G. Dawson, "Early behavioral intervention, brain plasticity, and the prevention of autism spectrum disorder," *Development and psychopathology*, vol. 20, no. 3, pp. 775–803, 2008.

[5] C. Lord, E. H. Cook, B. L. Leventhal, and D. G. Amaral, "Autism spectrum disorders," *Neuron*, vol. 28, no. 2, pp. 355–363, 2000.

[6] G. Dawson, A. N. Meltzoff, J. Osterling, J. Rinaldi, and E. Brown, "Children with autism fail to orient to naturally occurring social stimuli," *J. of autism and developmental disorders*, vol. 28, no. 6, pp. 479–485, 1998.

[7] K. A. Pelphrey, N. J. Sasson, J. S. Reznick, G. Paul, B. D. Goldman, and J. Piven, "Visual scanning of faces in autism," *J. of autism and developmental disorders*, vol. 32, no. 4, pp. 249–261, 2002.

[8] S. Wang, M. Jiang, X. M. Duchesne, E. A. Laugeson, D. P. Kennedy, R. Adolphs, and Q. Zhao, "Atypical visual saliency in autism spectrum disorder quantified through model-based eye tracking," *Neuron*, vol. 88, no. 3, pp. 604–616, 2015.

[9] B. M. Kuehn, "Cdc: autism spectrum disorders common," *Jama*, vol. 297, no. 9, pp. 940–940, 2007.

[10] T. Guha, Z. Yang, R. B. Grossman, and S. S. Narayanan, "A computational study of expressive facial dynamics in children with autism," *IEEE Trans. on Affective Computing*, vol. 9, no. 1, pp. 14–20, Jan 2018.

[11] P. R. Krishnappa Babu, P. Oza, and U. Lahiri, "Gaze-sensitive virtual reality based social communication platform for individuals with autism," *IEEE Trans. on Affective Computing*, vol. 9, no. 4, pp. 450–462, Oct 2018.

[12] T. Tang, "Helping neuro-typical individuals to "read" the emotion of children with autism spectrum disorder: an internet-of-things approach," in *The 15th Int'l Conf. on Interaction Design and Children*, 06 2016, pp. 666–671.

[13] L. Zhang, J. Wade, D. Bian, J. Fan, A. Swanson, A. Weitlauf, Z. Warren, and N. Sarkar, "Cognitive load measurement in a virtual reality-based driving system for autism intervention," *IEEE Trans. on Affective Computing*, vol. 8, no. 2, pp. 176–189, April 2017.

[14] J. Hashemi, G. Dawson, K. L. Carpenter, K. Campbell, Q. Qiu, S. Espinosa, S. Marsan, J. P. Baker, H. L. Egger, and G. Sapiro, "Computer vision analysis for quantification of autism risk behaviors," *IEEE Trans. on Affective Computing*, 2018.

[15] W. Liu, M. Li, and L. Yi, "Identifying children with autism spectrum disorder based on their face processing abnormality: A machine learning framework," *Autism Research*, vol. 9, no. 8, pp. 888–898, 2016.

[16] T. Nguyen, T. Duong, S. Venkatesh, and D. Phung, "Autism blogs: Expressed emotion, language styles and concerns in personal and community settings," *IEEE Trans. on Affective Computing*, vol. 6, no. 3, pp. 312–323, July 2015.

[17] S. Wang, S. Fan, B. Chen, S. Hakimi, L. K. Paul, Q. Zhao, and R. Adolphs, "Revealing the world of autism through the lens of a camera," *Current Biology*, vol. 26, no. 20, pp. R909–R910, 2016.

[18] O. Chum, J. Philbin, A. Zisserman *et al.*, "Near duplicate image detection: min-hash and tf-idf weighting." in *Bmvc*, vol. 810, 2008, pp. 812–815.

[19] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *Int'l J. of Computer Vision*, vol. 42, no. 3, pp. 145–175, 2001.

[20] J. M. Mateos-Pérez, R. Redondo, R. Nava, J. C. Valdiviezo, G. Cristóbal, B. Escalante-Ramírez, M. J. Ruiz-Serrano, J. Pascau, and M. Desco, "Comparative evaluation of autofocus algorithms for a real-time system for automatic detection of mycobacterium tuberculosis," *Cytometry Part A*, vol. 81, no. 3, pp. 213–221, 2012.

[21] S. Helgason and S. Helgason, *The radon transform*. Springer, 1999.

[22] R. N. Bracewell and R. N. Bracewell, *The Fourier transform and its applications*. McGraw-Hill New York, 1986, vol. 31999.

[23] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields," in *arXiv:1812.08008*, 2018.

[24] C.-C. Chang and C.-J. Lin, "Libsvm: A library for support vector machines," *ACM trans. on intelligent systems and technology*, vol. 2, no. 3, p. 27, 2011.

[25] V. N. Vapnik, *Statistical learning theory*. Wiley, 1998.

[26] V. Hus and C. Lord, "The autism diagnostic observation schedule, module 4: revised algorithm and standardized severity scores," *Journal of autism and developmental disorders*, vol. 44, no. 8, pp. 1996–2012, 2014.

[27] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, p. 436, 2015.

[28] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.