Automatic Analysis of Eye Tracking Data for Medical Diagnosis

Filippo Galgani, Yiwen Sun, Pier Luca Lanzi, Jason Leigh

***Abstract*— Several studies have analyzed the link between mental dysfunctions and eye movements, using eye tracking techniques to determine where a person is looking, that is, the fixations. In this paper, we present a novel methodology to improve current diagnosis and evaluation methods of attention disorders. We have developed and tested several data-mining methodologies suitable for the automatic analysis and visualization of eye tracking data. In particular three novel methods of classification of subjects are proposed: (i) a method that uses Expectation Maximization to classify according to statistical likelihood of fixations locations; (ii) a procedure based on the Levenshtein distance method to compare sequences of fixations; and (iii) a method based on the analysis of the transitions frequencies of fixations between regions. Results of evaluation of classification accuracy are finally presented.**

I. Introduction

THROUGH the use of eye tracking techniques an individual's eye movements are measured so that it is possible to know where a person is looking at any given time and the sequence in which the eyes are shifting from one location to another. The eye movements as a reflection of cognitive processes have been investigated in the field of psychology for over fifty years, with many studies supporting the view that shifts in viewer attention are reflected by changes in the point of visual fixation. Accordingly eye tracking is widely accepted as a valuable method for studying the visual attention and cognitive state of a subject (for a review, see [1]). Numerous studies have used eye movements as diagnostic tool contributing to an increased understanding of the neurological basis of many attention disorders, such as Attention Deficit Hyperactivity Disorder (ADHD), Autism, Dementia, Stress Disorder and others.

We collected and analyzed the eye movements of a group of subjects, in a free image viewing task, in which each subject was presented with a sequence of images, and left free to look at them without being given any instruction. In this work, our hypothesis was that in this kind of task the eye movements of subjects with attention disorders would significantly differ from the eye movements of control subjects who were not diagnosed with any mental disorder. To our best knowledge, there is no previous study that verifies the behavior of individuals affected by ADHD in this kind of task. For this purpose, we have studied, developed and tested several data-mining methodologies suitable for the automatic analysis and visualization of eye tracking data, including finding convenient ways to visualize different aspects of the picture scanning, to highlight similarities and differences between the behavior of different subjects, and ultimately to perform automatic classification of subjects.

Filippo Galgani and Pier Luca Lanzi: Politecnico di Milano, Italy
Yiwen Sun and Jason Leigh: University of Illinois at Chicago

We present three novel methods of classification: a method that uses Expectation Maximization [2] to model the distribution of fixations over an image and classify according to statistical likelihood of fixations' locations; a procedure based on the Levenshtein distance method [3] to compare sequences of fixations; a method based on the analysis of the transitions frequencies of fixations between regions of the image. With eye tracking data recorded on a population of both control subjects and diagnosed with mental disorder ones, we evaluated the accuracies of the three proposed methods.

The rest of this paper is organized as follows: Section II describes the task addressed by our work; Section III presents an overview of the related work; Section IV describes the way we gathered eye movement data; Section V, VI and VII present each one of the three methods presented for classification. In Section VIII, we draw the conclusions.

II. Problem Description

Eye movements bring the fovea (that is, the area centered on the gaze point that is seen in high detail) to Regions of Interest (ROIs) within a scene to be further examined. These inspections are fixations interspersed by rapid eye movements, called saccades. A sequence of fixations is defined as a scanpath. It can be assumed that visual attention follows the fovea. Although this is not always the case (one can attend to an object in their periphery), non foveal visual attention is immeasurable and unlikely in most tasks. Consequently scanpath data can be examined to determine what a subject attends to and thus where the ROIs are located within a scene. Eye tracking data usually come in the form of a sequence of temporized samples that represent the location of the gaze of the subject over a certain stimulus, the process of extracting useful knowledge from this kind of data is definitely not trivial, and a complex analysis is often required. The objective of this work is to test the hypothesis that, when presented with a certain kind of stimuli, the eye movements of subjects who suffer from attention disorders will systematically differ from the eye movements of control subjects who are not diagnosed with any mental disorder. Our approach consisted in recording eye-tracking data on a population of control subjects and subjects with a mental disorder such as autism or attention deficit. Then we applied data-mining techniques, to identify a base line for normal behavior and discover characteristics to distinguish whether a patient is suffering from a mental disorder. We showed a sequence of images to each subject, the images were of different kinds (photos, drawings, geometrical shapes, etc.). Subjects were free to look at them in any way, without being given any instruction. The only constraint in their

visual exploring was the limited time during which each image was displayed. In contrast to what have been the most common approach when investigating the relationship between eye movements and mental dysfunctions, we do not have a particular task (for example the prosaccade or antisaccade paradigm[4]) that the subject is asked to perform, and a particular series of measure (for example saccadic reaction time) to test the differences with a control group. In our case there is no predefined right or wrong behavior, as there is no direct definition of performance or errors.

The methodologies that we present characterize the data and automatically identify subjects affected by attention disorders. For each subject, we have five seconds data per image, given by the location of the gaze (x-y coordinates in pixels over the screen) and the timing information (timestamp and image index). An example of the raw data we start with is given in Figure 1, where the gaze points of two subjects are super-imposed onto the corresponding image.

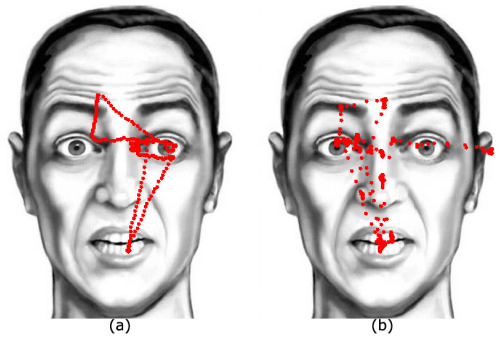

Fig. 1. Eye movement for a single image for a five seconds lapse of time: (a) a control subject; (b) a subject diagnosed with ADHD.

Our data consist of subjects divided in two classes: one class is formed by the control subjects (subjects not diagnosed with any disorder, class C), the other one consists of all subjects diagnosed with attention disorders (class A). We consider our set of data as the baseline for defining subjects' behavior in our image viewing task, and we are looking for a model that generalizes the class description from the training subjects, and be able to classify any novel subjects.

III. Related Work

Many efforts have been aimed at explaining, and in this way being able to predict, where the gaze is more likely to be directed in different conditions. Eyes are naturally directed on the salient or important areas. In some structured task it is easier to recognize common patterns of eye movement, for example, left to right, top to bottom patterns in reading process. While in free scene viewing it is more difficult to identify apparent strategies. It is usually recognized that global information about scene background or setting is extracted during the initial fixation, the gist of the scene is abstracted on the first few fixations, and the remainder of fixations are used to fill in details, in general if an object is important it is usually fixated. However there is no simple way of telling what the brain is doing during a particular visual scan of the scene, because it is possible to visually fixate one location while simultaneously diverting attention to another, examining a scanpath over a visual stimulus let us identify only which specific regions were looked at, but we can not be sure on where the attention of the subject was directed. Two general hypotheses have been advanced to explain fixation locations in scenes. According to the first one, the visual saliency hypothesis, fixation sites are selected based on image properties generated in a bottom-up manner from the current scene. In contrast, according to what is called the cognitive control hypothesis, fixation sites are selected based on the needs of the cognitive system in relation to the current task. Most researchers agree that eye movement targeting involves a combination of bottom up and top down guidance factors [5].

Eye tracking tasks have been widely used as biological markers of disease because they are objective, painless, and non-invasive, and they can provide insight into the neural substrate underlying the pathophysiology. Among attention related disorders we are mostly interested in Attention Deficit Hyperactivity Disorder (ADHD). This disorder is characterized by the symptoms of impulsiveness, hyperactivity, and inattention, it is a prevalent neurobehavioral disorder estimated to affect 5% of children and, for some, these core symptoms are believed to persist into adulthood. Children with ADHD often perform poorly on tasks requiring the sustained and systematic allocation of attention over periods of extended time. Also response inhibition is considered an important component of the disability because ADHD subjects have difficulty suppressing inappropriate behavioral responses. There exist many eye tracking based studies which explore the eye movements of children and adults diagnosed with ADHD and controls recruited to perform a series of saccadic eye movement tasks, those studies have contributed to an increased understanding of the neurological basis of many attention disorders. Different saccade tasks are employed (prosaccades and antisaccades [6][4] , memory-guided saccades [7] , countermanding paradigm [8], attention blink paradigm [9]),to examine functions necessary for the planning and the execution of eye movements, including motor response preparation, response inhibition, and working memory. The subject is usually asked to perform simple eye movements towards (or in the opposite direction to) markers, to react to sounds or to follow stable or moving dots [10]. The two main characteristics that distinguish ADHD subjects from control ones are the difficulty to voluntarily suppress inappropriate behavioral responses (reflexive saccades), and the problems in maintaining steady fixation.

IV. Data Collection

In collaboration with a team of clinical psychologists at Visionary Sciences, we have designed several stimuli to be employed in our experiments. The stimulus we used to validate the procedures described in this paper consists of an ordered and temporized sequence of images, each of which is displayed for 5 seconds to the subject, for a total duration of 125 seconds.

We recorded eye tracking data from two different groups of subjects: a group of controls and a group of patients diagnosed with attention disorders. The group of controls consist of eighteen volunteer subjects, which were recruited among the students and staff in college, where ISCAN Polhemus VisionTrak Binocular Desktop 300 System [11] was used to record eye movement. None of the subjects in this group is diagnosed with attention disorders. All the subjects had normal, or corrected to normal, vision, and some of them took the test with eyeglasses or lenses. The group of patients diagnosed with attention disorders were tested in a specialized clinic, where an Eyegaze System from LC. Technologies Inc [12] was used. Most subjects in this group are diagnosed with ADHD, two with ADHD and Bipolar Disorder and two others with Post Traumatic Stress Disorder.The age of patients ranges from 9 to 59 years old.

The raw data obtained from the eye tracker consist of sequences of samples, given by the Cartesian coordinates and timestamps. Raw eye-movement data points are extracted to fixations using Dispersion-Threshold Identification (DTI) [13] algorithm. Then we applied an higher level clustering of fixations to individuate Regions of Interest (ROIs) on the image. These automatically identified ROIs will be used in the automatic classification of the data described in the following Sections.

V. EM Approach

The first approach we present for the automatic classification of subjects is based on Expectation Maximization [2] (EM). EM is commonly used as a clustering algorithm [14]. We used it to model the distribution of fixations over the image. The idea underlying the EM approach is to build a model that characterizes the two groups of subjects, identifying a distribution of fixations over the image for the control behavior and one for the attention disorder behavior, so that it is possible to determine if any new subject's data is more likely to have been generated by one or the other distribution. Starting with a trace of the recorded eye movements over time for a certain number of images, and for the moment considering the case of a single image, we put all fixations together not considering the timestamps. We separated the two groups of data files, C (control subjects) and A (subjects with ADHD). For each of these two groups we computed two set of clusters using the EM algorithm, aggregating all the data from control subjects relative to the considered picture, and processing all points together, then repeating the same process with subjects diagnosed with disorders. We have two sets of clusters: set SC (from subjects labeled as C) and set SA (from subjects with attention disorder, labeled as A). Before applying the clustering algorithm itself, the data is firstly parsed into fixations using the dispersion threshold identification algorithm [13], meanwhile saccades are removed. Then noise points are individuated and removed using the meanshift clustering algorithms. To classify a previously unseen subject, we compare his/her distribution of fixations in the current image to the two sets of clusters SC and SA, in order to establish which one of the two distributions is more likely to have generated the data. Each of the two set of clusters (SC and SA) obtained with the EM algorithm is formed by a certain number of clusters. The clusters are modeled as Gaussian distributions, and identified by their prior probabilities ($Pr(j)$) and two dimensional means ($\mu_j = \left(\mu_j^x, \mu_j^y\right)$) and variances ($\sigma_j = \left(\sigma_j^x, \sigma_j^y\right)$) . Given the new set of fixations, we estimate the probability of the data given the set of clusters SC and given the set of clusters SA, and check which one gives an higher log-likelihood. Let $p_i = (x_i, y_i)$ be each fixation of the dataset, and given the density function of each cluster C_j (the normal bivariate distribution):

$$Pr(p \mid C_j) = \frac{1}{2\pi\sigma_j^x\sigma_j^y} \cdot \exp\left(-\frac{(\mu_j^x - x)^2}{2(\sigma_j^x)^2} - \frac{(\mu_j^y - y)^2}{2(\sigma_j^y)^2}\right) \quad (1)$$

Using the following equation, that is used in the EM clustering process during the Estimation (E) step, we can calculate the total loglikelihood with the following equation:

$$\begin{aligned} loglikelihood &= \log \prod_i Pr(p_i) = \sum_i \log Pr(p_i) \\ &= \sum_i \log\left(\sum_j Pr(j) \cdot Pr(p_i \mid C_j)\right) \\ &= \sum_i Pr(j) \frac{1}{2\pi\sigma_j^x\sigma_j^y} \cdot \\ &\cdot \exp\left(-\frac{(\mu_j^x - x)^2}{2(\sigma_j^x)^2} - \frac{(\mu_j^y - y)^2}{2(\sigma_j^y)^2}\right) \quad (2) \end{aligned}$$

We compute the total log-likelihood with (2), and compare the result for the two clusters sets SC and SA, if the log-likelihood for set SC is higher than the log- likelihood for SA, then we classify the subject as control (label C), otherwise we classify the subject as affected by a disorder (label A). For an example see Figure 2. The classification algorithm is

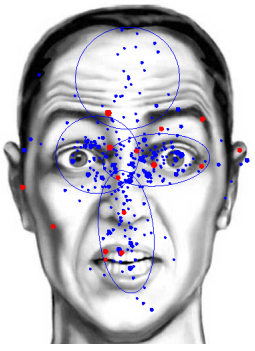

Fig. 2. Data of a single subject (in red) is compared to the cluster set resulting from the group of control subject (in blue).

the following:

Model building:

1) Take the training set, divide it in two sets, C and A, according to the class labels.

2) Parse the fixations of each trace (with the DTI algorithm).
3) Remove the noise from each trace using the meanshift clustering algorithm.
4) Compute and store a set of clusters for C (SC), and a set for A (SA), using the EM clustering algorithm.

Classification of new instances:

1) Take the new instance I.
2) Parse the fixations of I, and remove noise points.
3) Calculate the likelihood Lc (likelihood of all the fixations given the set of cluster SC) and La (likelihood with respect to the cluster set SA) of all the fixations using (2).
4) If $Lc < La$ classify I as C, if $La > Lc$ classify I as A

With the above equations and algorithm the same process is applied to all data points (all the fixations made by the subject to be classified). Another way to proceed is to apply the EM clustering algorithm to the fixations of the subject to be classified, compute the clusters, and then consider only the resulting centroids in relation to the two sets of clusters. The procedure is the same as above, but in (2) the index i of the summation ranges only on the centroids obtained, not on all points.

With the method just described we obtain a classifier for each of the images in a sequence. Because a subject normally goes through all the package, we can improve classification performance if we find a way to combine the predictions relative to each image. To obtain a single classification response we have to ensemble the single responses into a final unique label, we do that by having each image to vote in a binary fashion either for class label C or A, according to which class gives higher likelihood, and then counting the total number of votes to obtain a majority response.

Counting the number of errors (or misclassified subjects) on a test set, we measured the classification error as the percentage of misclassified subjects. The classification error is our performance measure, the lower is the error, the more accurate is the model. To form training and test sets we use the bootstrap technique [15], which is particularly suitable for small dataset as ours, using sampling with replacement to form ten training and test sets, and then averaging the results over the ten iterations. Expectation Maximization clustering requires to specify the number of desired clusters, for this reason we have used different values: we have tested the algorithm using three and four clusters for each image. Then, instead of keeping fixed number of clusters for all images, we have used a cross validation procedure to establish image by image the most appropriate number of clusters (the one that maximize the likelihood), following the approach used in the WEKA tool [14]. We have tried both to classify subjects according to the likelihood of all points, or only of the centroids computed with the EM clustering (as described previously). The obtained results can be found in Table I.

In machine learning, a common method to improve the performance of a classifier consists in combining the output of different models, and several techniques exist to do it by learning an ensemble of methods and using them in combination. Among these, we used the Bagging technique [16]. Bagging is based on the idea of creating several datasets by sampling with replacement from the original dataset, building a model for each of those, and at the end combining the various output into a single prediction. We compared the performance of our method using bagging to the standard version(computing three clusters and using all fixations), the result can be found in Table II.

TABLE I
CLASSIFICATION ERROR OF THE EM METHOD.

Type of Learning	$\mu \pm \sigma$
Using 3 cluster, considering all the fixations	0.273±0.112
Using 3 cluster, considering only the centroids	0.309±0.120
Using 4 cluster, considering all the fixations	0.275±0.146
Using 4 cluster, considering only the centroids	0.328±0.133
Variable number of cluster, considering all the fixations	0.286±0.101
Variable number of cluster, considering only the centroids	0.381±0.095

TABLE II
CLASSIFICATION ERROR OF THE EM METHOD WITH BAGGING.

Type of Learning	$\mu \pm \sigma$
Using bagging	0.368±0.126
Standard learning	0.339±0.104

VI. LEVENSTHEIN APPROACH

There is a certain amount of difficulty in quantifying the similarity between two scanpaths. This difficulty lies in condensing the spatial information of multiple fixations without losing the sequence information inherent in a scanpath. The most popular technique for quantifying the similarity of such sequences is the Levenshtein, or string-edit, distance. This method to perform pair-wise scanpath comparisons uses ROI-based alphabet encoding: the sequence of fixations is translated into a sequence of symbols that identifies the areas where the fixations were posed, by assigning symbols to different ROIs. We already described how using the meanshift clustering algorithm we can find automatically Regions of Interest in the image, assign each fixation to an area, and identify noise (fixations that are far from any region). We concatenate the cluster labels assigned by the clustering algorithm to the fixations of a subject and obtain a string describing the scanpath. If we are working with a single subject, we apply the clustering algorithm to the set of fixations, if we want to compare two sets of fixations from two subjects, we merge the two sets of points and apply the clustering algorithm to the resulting total set of data points, in this way we find the overall regions of interest, some of which will be attended by both subjects, while other will contain fixations only from one subject. The same process can be applied to more than two subjects. It is possible to obtain a compressed scanpath string, removing the equal consecutive symbols, in this way we keep track of which

regions the subject attended, but not how many fixations he posed on each of them. Fixations that are identified as noise, being out of any cluster, can be inserted with a special index (for example 0) or can be removed from the scanpath. This representation of the fixation sequence approximates the information on the exact spatial locations of fixations (in term of pixels coordinates), substituted with the ROI to which it is assigned, but it is more convenient for the task of comparing these sequences. An example of two scanpath strings can be found in Figure 3. A standard approach to compare two such

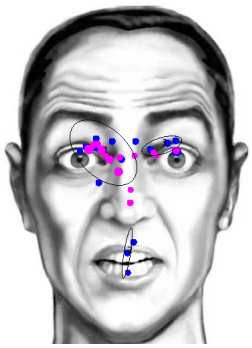

Fig. 3. An example of reduction of a set of fixations to a character sequence representation. The first scanpath string is 11112313122113113, the second 11111333111111.

sequences consists in analyzing the total difference between two sequences into a collection of individual elementary differences, the distance of the two sequences can be seen as the amount of elementary differences that distinguishes one sequence from the other. It is convenient to treat the elementary differences as elementary operations, and to think of the operations as actively changing a source sequence into a target sequence, step by step. We consider as elementary operations substitutions, deletions, and insertions (the word indel can be used to indicate both insertion and deletion). The process of finding a sequence of edit operations that gives one string from the other is called sequence alignment, where the alignment itself is the sequence of operation (substitutions and indels), and the similarity between two strings is computed by calculating the minimum number of editing steps required to turn one sequence into the other.

The concept of distance from one sequence to another, taken as the smallest number of substitutions and indels required to change the first into the second, was introduced by Levenshtein [3]. An algorithm for calculating the minimum editing cost, with application to eye tracking data, is given in [17]. Privitera and Stark [18] develop a methodology that uses string edit techniques to compare scanpaths, in order to compare artificially created scanpaths with the ones recorded by humans, and thus assess the performance of the algorithms that try to predict which points are more likely to be fixated. A similar analysis is employed in [19]. The string edit methodology has been since then applied in several studies. Josephson and Holmes evaluated web pages browsing [20] and television watching behavior [21]. Their experimental results showed that while some individuals displayed scanpaths that resembles each other over time, in many instances the most similar sequences were from different subjects. In both of their studies, the viewing stimulus was manually partitioned into Regions Of Interest a priori, thus precluding the need for automatic cluster analysis. They modified the string editing comparison by introducing an edit cost (that was kept constant to one by Privitera and Stark), which is based on adjacency of regions (with adjacent regions having lower substitution cost). Takeuchi and Habuchi [22] tried to improve the algorithm using Euclidean distance between centroids of the regions as substitution cost (empirically finding that this method gives better results), they still kept deletion and insertion cost unitary. Heminghous, [23] and [24], proposed the use of the meanshift clustering algorithm [25] to detect automatically the ROIs (regions of interest), noting several advantages of this clustering algorithm, that is particularly adapted to eye movement data analysis and in this way permit to find an automatic method of detecting the meaningful areas of a picture. Hembrooke et al. [26] explored the possibility of finding a single, representative average scanpath, using string-editing as a multiple sequence alignment algorithm between several traces. Recently, in an application that implements also the Needleman-Wunsch algorithm [27], allowing the users to specify scoring parameters for the alignment, West et al. [28] extended the string-editing scanpath comparison approach with the use of a local alignment algorithm [29], for finding common (or closely related) subsequences, which do not need to be located in the same position in both sequences, in this way visual strategy shared by two individuals can be identified, even if they occur at different times within an experiment.

We used as dissimilarity measure between scanpaths the overall cost of the operations needed to get one string from the other. The algorithm for finding the minimum cost of alignment that we refer to is the one developed by Levenshtein [3], enhanced with the Needleman-Wunsch approach [27] to let the user specify different cost functions. The Levenshtein algorithm works using a dynamic programming approach: let A and B be the strings to be compared, n the length of A and m the length of B, A[i] the i-th element of A, A[1..i] the first i elements of A (and the same for B). The key observation for the alignment problem is that the edit cost between the sequences A[1..n] and B[1..m] can be computed by taking the minimum of the three following values:

- The cost of aligning A[1..n-1] and B[1..m-1], plus the scoring of substituting A[n] with B[m].
- The cost of aligning A[1..n-1] and B[1..m] plus the scoring of deleting A[n].
- The cost of aligning A[1..n] and B[1..m-1] plus the scoring of inserting B[m].

The algorithm builds a matrix of the dissimilarity score for every two subsequences A[1..i] and B[1..j], starting from the base case cost(A[0],B[0])=0, where A[0] and B[0] are defined as empty strings. In the Levenshtein algorithm all the costs are kept fixed to one. We normalize the dissimilarity value of the two sequences by the average of the two lengths of the sequences, otherwise longer scanpaths would

TABLE III
CLASSIFICATION ERROR OF THE LEVENSHTEIN METHOD.

Type of Learning	$\mu \pm \sigma$
Unitary substitutions costs, considering the full scanpath	0.268±0.071
Unitary substitutions costs, considering the compressed scanpath	0.153±0.082
Substitutions costs proportional to distance, considering the full scanpath	0.361±0.062
Substitutions costs proportional to distance, considering the compressed scanpath	0.118±0.058

be penalized. This makes the distance relative to length and comparable across pairs of varying lengths. While in the original Levenshtein algorithm all edit costs are fixed to one, in our implementation it is possible to define a similarity matrix to specify the score for each mismatch between a pair of characters in the alignment. Deriving a similarity measurement from flexible parameters is desirable in situations in which two different ROIs have a close spatial proximity or serve a similar function. In these situations, we might expect the two ROIs to be interchangeable in some fixation sequences. To have substitutions of spatially close gaze points costing little, and substitution between long distance points costing more, we compute the centroids of all the regions, and set the substitution costs directly proportional to the distances between the centroids of the corresponding regions, as it is suggested in [22].

With the procedure just described, given any two traces, we can obtain a numeric index of similarity between the two. We used this similarity definition to perform automatic classification of subjects. The more intuitive way to classify subjects according to this similarity measure is to perform a comparison of the subject to be classified against a population of control and diagnosed subjects and compare the resulting similarities. The Levenshtein distance is defined only between a pair of traces, and it cannot be applied to individuals against a group. For this reason, to classify a new instance, we compare its trace with all the traces in the C group, one by one, and calculate the average similarity. We then do the same with the A group, and obtain another average similarity, that estimates the similarity between the new instance and the set A. The new subject is classified according to which of the two similarity values is higher. We build one model for each image, and then combine the response of each image in a final label using majority voting.

We estimated the performance of this method using the bootstrap method. In the tests we used both unitary substitution costs and set the substitution costs proportional to the distances between ROIs to which the symbols refer. Both the extended and compressed version of the scanpath are used. Results are presented in Table III.

VII. TRANSITIONS ANALYSIS APPROACH

The third classification approach we describe is based on the analysis of transitions between ROIs. Stark and Ellis [30] derived Markov matrices from the letter strings, demonstrating the existence of a few structured processes. Ellis and Smith [31] elaborated on Noton and Stark's scanpath phenomenon [32] by claiming that scanpaths can be generated by completely random, stratified random, or statistically dependent stochastic processes, but they did not test these conjectures. Ellis and Stark [33] tested the hypothesis that scanpaths are statistically dependent, and not random, in visual information seeking in dynamic visual environment. They investigated the statistical dependency of fixation sequences by looking at the frequencies of transitions between each ROI in a stimulus and performed a chi-square goodness-of-fit test of transition frequencies found in subjects fixation sequences against those that would result from a stratified random sampling of the data, and determined that the frequency of a transition to a given ROI is statistically dependent upon the previous ROI: even if the resultant statistical dependency is low in magnitude, random sampling models do not completely account for the observed patterns. Hacisalihzade, Stark and Allen [34] investigated the possibility of modelling sequences of fixations as Markov processes, generating Markov matrices of transitions of different order, and then quantifying the similarity of eye movements comparing them with an absolute error index (difference between the matrices) and string editing methodologies. Pieters et al. [35] in a study about repeated exposure to printed advertising, test and validate the hypothesis that scanpaths belong to a stationary, reversible, first order Markov process.

Dynamic analysis of the eye movement can be modeled as stochastic process, where the ROIs correspond to states, and the eye movements between these ROIs to state transitions. We define the location of the gaze at sample n as a random variable Ln, $\{L_n, n \in N\}$ is the stochastic process, where the index set N is the set of all samples. At a particular time, the system is found to be in exactly one of the states, that we label $0, 1, , S$, each state is one of the ROIs computed with the clustering algorithm, an additional state is added to represent all other areas not encompassed by the clusters (state 0), for the points labeled as noise, so that the states are exhaustive. The conditional probabilities $P\{L_{n+1} = j \mid L_n = i\}$, called transition probabilities, denote the probability that the system will be in state j given that it was last in state i. We can compute the matrix of transition frequency, where the entry ij of the matrix represent how many times a fixation in the i-th ROI was followed by one fixation in the j-th region. We compute the ROIs of the image with the application of a clustering algorithm, and count the transitions of the gaze between the different ROIs. We calculate probabilities from the observed frequencies, normalizing each row so that the sum of its elements is equal to one. Note that a matrix of transition probabilities can be used to represent the behavior of a single subject, but as well of a group of subjects, in a straightforward way: counting all the transitions made by the group of subjects. Given the transition matrices, we have several ways to employ these matrices to compare the behavior of the subjects. First of all we can compute matrices of a single user, as well as a group of users, and compare individual matrices against one or more groups. Hacisalihzade, Stark and Allen [34] describe a way to obtain

TABLE IV

CLASSIFICATION ERROR OF THE TRANSITIONS ANALYSIS METHOD.

Type of Learning	$\mu \pm \sigma$
Using absolute mean error, considering the full scanpath	0.204±0.075
Using absolute mean error, considering the compressed scanpath	0.335±0.137
Using correlation, considering the full scanpath	0.165±0.063
Using correlation, considering the compressed scanpath	0.304±0.113

a numerical index of similarity between two subjects (or group of subjects), given two transition matrices $M1$ and $M2$. They define the error, or statistical discordance, between the two observed matrices as: $E = M1 - M2$. A possible scalar measure of the statistical discordance matrix E is the typical error of each element defined as:

$$\bar{e} = \frac{1}{n^2} \sum_{i,j=1}^{n} |e_{ij}| \tag{3}$$

where e_{ij} is the ij-th element of the error matrix E and n its dimension. Another way to obtain a similarity measure is to calculate the Pearson product-moment correlation between the two matrices. This method has been used by Ellis and Stark [33].

The main idea of classification method based on transition analysis is to compare matrices of transitions between different ROIs of the subject to be classified, with those of the C and A groups of subjects, to see which group it is more similar to. We performed the comparison in the same way we did with Levenshtein method, making many pairwise comparisons between single subjects, and then taking the average of the similarities. Another way of proceeding is to calculate the matrices of transition probabilities for the whole C and A groups, and then compare the matrix of the new subject to these two, obtaining directly two similarity values, we tried both methods but the obtained results for the latter method was worse. Once we obtain a similarity coefficient between the matrices, we can classify the subject according to which group (C or A) gets the higher value of similarity. For similarity measure we can use both the absolute mean error (which actually is a dissimilarity, so we use $1 - err$) and the correlation between the coefficients of the matrices. As done with the other two classification methods, we performed an evaluation of performance with the bootstrap technique. In the test we used both the normal and compressed representation of the scanpaths. The result can be found in Table IV.

VIII. CONCLUSIONS

We have described three original methodologies to perform the task of automatic classification of subjects into the two classes that we identified. The first method models the distribution of fixations over each image using Expectation Maximization clustering, and for any new subject computes the likelihood of his distribution of fixation with respect to those of the group of control subjects and of the group of subjects diagnosed with ADHD, the subject is then classified according to which of these two likelihood values is higher. The second method is based on the Levenshtein distance, it computes the average similarity of the behavior of the subject to be classified with the two groups of subjects, to estimate how similar his viewing behavior is with the other control or ADHD subjects. The third method is based on the analysis of matrices of transition probabilities between the ROIs of the image, and uses two coefficients of similarity (absolute error and correlation) to estimate how similar the transition behavior of the subject is to the two groups of subjects.

We tested the effectiveness of our methods, using a database containing the data of 43 individuals relative to a specific sequence of images. 18 individuals are control subject, while 25 have a diagnosis of an attention disorder. A tool has been developed and implemented in C++, and was used to evaluate the performances using bootstrap [15]. The best result is obtained with the Levenshtein distance method, with an average error of the 11.8%, the best result obtained with the transitions analysis method is slightly worse, with an error of the 16.5%, Expectation Maximization performs worst with an error of the 27.3%.

It is difficult to evaluate globally the goodness of these results, as they are pioneer, and to our knowledge there is no previous work that attempts to classify subjects in relation to attention disorders, based on the subjects' eye movements. For this reason it is not easy to understand the impact of the results. Previous work on the analysis of eye movements of subjects with ADHD was based on much simpler tasks, like the pro-saccade, anti-saccade, go-stop tasks, and measured specific aspects of the eye movement, as saccadic reaction time, precision, directional errors, to establish if a statistically significant deviation in some of these metrics was present in individuals with ADHD. None of these studies, however, was using supervising learning to automatically classify unclassified subjects as we did. The task that we used is also different, and more difficult to be analyzed, from those studies, because we present stimuli to the subjects without giving any instructions or having an expected behavior, thus there is not one metric to be measured. One of the objectives of this study was to explore the hypothesis that in this kind of task the eye movement of subjects who suffer from attention disorders would differ systematically from those of control subjects. The results obtained show that this difference exists, although it is not easily quantifiable. From a qualitative point of view, the fact that EM perform the worst can be explained considering the fact that EM does not take into consideration the dynamic nature of the fixations, but only their locations over each image, while the Levenshtein distance and the transitions analysis methods consider also the order in which these fixations are made.

As all the studies involving human subjects, the main limitation to this work is the restricted size of the available data. A bigger population of subjects would help in estimation of the effectiveness of our algorithms, and would produce a better estimation of their accuracies. We also have to consider that in dealing with the population of subjects

we have introduced a simplification: we have divided all the subjects in two classes, the control subjects and the subjects diagnosed with mental disorders. While the class of control subjects can be considered quite homogeneous, the main factor of difference being the age of the individuals, the class of subjects with attention disorders is composed of heterogeneous type of subjects, with differences in age, clinical situation and medication treatment. It would be important to test if our methods can help in monitoring a clinical situation, highlighting the changes in behavior of one subject, and helping the physicians to understand not only if a patient is suffering a certain disorder, but also describing the clinical evolution of the subject. For this reason, classification accuracy, as we measured, can be limiting for describing the actual usefulness of the algorithms. A better way to check their effectiveness in describing a patient's situation would be to compare the numerical values, the coefficients that the algorithms output, to a qualitative description of the state of the patient made by a psychologist, and this has not been possible so far. Other possible directions include, (i) the evaluation on the quality and appropriateness of the stimuli employed in the experiment and (ii) the extension of the analysis by incorporating some spatial features of the images involved.

ACKNOWLEDGMENTS

We would like to thank Drs Richard Markin, Len Carr and Anthony Vertino of Visionary Sciences for providing clinical expertise for this project.

REFERENCES

[1] K. Rayner, "Eye movements and information processing: 20 years of research," *Psychological Bulletin*, vol. 124, no. 3, pp. 372–422, 1998.

[2] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistical Society, Series B*, vol. 39, no. 1, pp. 1–38, 1977.

[3] V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions and reversals," *Doklady Physics*, vol. 10, pp. 707–710, 1966.

[4] D. Munoz, I. Armstrong, and B. Coe, "Using eye movements to probe development and dysfunction," *Eye Movements: A Window on Mind and Brain*, 2007.

[5] J. Henderson, J. Brockmole, M. Castelhano, and M. Mack, "Visual saliency does not account for eye movements during visual search in real-world scenes," *Eye Movements: A Window on Mind and Brain*, pp. 537–562, 2007.

[6] D. Munoz, I. Armstrong, K. Hampton, and K. Moore, "Altered control of visual fixation and saccadic eye movements in attention-deficit hyperactivity disorder," *Journal of Neurophysiology, Physiological Soc.*, 2003.

[7] S. H. Mostofsky, A. G. Lasker, H. S. Singer, M. B. Denckla, and D. S. Zee, "Oculomotor abnormalities in boys with tourette syndrome with and without adhd," *Journal of the American Academy of Child and Adolescent Psychiatry*, vol. 40, no. 12, 2001.

[8] I. Armstrong and D. Munoz, "Inhibitory control of eye movements during oculomotor countermanding in adults with attention-deficit hyperactivity disorder," *Experimental Brain Research, Springer*, 2003.

[9] ——, "Attentional blink in adults with attention-deficit hyperactivity disorder," *Experimental Brain Research, Springer*, 2003.

[10] T. Gould, T. Bastain, M. Israel, D. Hommer, and F. Castellanos, "Altered performance on an ocular fixation task in attention-deficit/hyperactivity disorder," *Biological Psychiatry*, vol. 50, no. 8, pp. 633 – 635, 2001.

[11] [Online]. Available: http://www.polhemus.com

[12] [Online]. Available: http://www.eyegaze.com/index.htm

[13] D. Salvucci and J. Goldberg, "Identifying fixations and saccades in eye-tracking protocols," in *Proceedings of the Eye Tracking Research and Applications Symposium. Palm Beach Gardens, FL.*, 2000.

[14] I. H. Witten and E. Frank, *Data Mining: Practical machine learning tools and techniques*, 2nd ed. San Francisco: Morgan Kaufmann, 2005.

[15] B. Efron and R. Tibshirani, "An introduction to the bootstrap," *London: Chapman and Hall*, 1993.

[16] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.

[17] S. Brandt and L. Stark, "Spontaneous eye movements during visual imagery reflect the content of the visual scene," *Journal of Cognitive Neuroscience*, vol. 9, no. 1, pp. 27–38, 1997.

[18] C. M. Privitera and L. W. Stark, "Algorithms for defining visual regions-of-interest: Comparison with eye fixations," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 22, no. 9, pp. 970–982, 2000.

[19] ——, "Evaluating image processing algorithms that predict regions of interest," *Pattern Recognition Letters*, vol. 19, no. 11, pp. 1037–1043, 1998.

[20] S. Josephson and M. Holmes, "Visual attention to repeated internet images: Testing the scanpath theory on the world wide web," *Eye Tracking Research and Applications (ETRA) Symposium. ACM, New Orleans, LA.*, pp. 43–49, 2002.

[21] ——, "Clutter or content? how on-screen enhancements affect how tv viewers scan and what they learn," *Eye Tracking Research and Applications (ETRA) Symposium. ACM, San Diego, CA*, pp. 155–162, 2006.

[22] H. Takeuchi and Y. Habuchi, "A quantitative method for analyzing scan path data obtained by eye tracker," *Computational Intelligence and Data Mining, 2007. CIDM 2007. IEEE Symposium on*, pp. 283–286, 2007.

[23] Heminghous, "Icomp: A scanpath comparison tool," Master's thesis, 2006.

[24] Heminghous and Duchowski, "Icomp: A tool for scanpath visualization and comparison," *Proceedings of the Symposium on Applied Perception in Graphics and Visualization (APGV), Boston, MA*, 2006.

[25] A. Santella and D. DeCarlo, "Robust clustering of eye movement recordings for quantification of visual interest," *Eye Tracking Research and Applications (ETRA) Symposium. ACM, San Antonio, TX*, pp. 27–34, 2004.

[26] H. Hembrooke, M. Feusner, and G. Gay, "Averaging scan patterns and what they can tell us," *Eye Tracking Research and Applications (ETRA) Symposium. ACM, San Diego, CA*, p. 41, 2006.

[27] S. B. Needleman and C. D. Wunsch, "A general method applicable to the search for similarities in the amino acid sequence of two proteins," *Journal of Molecular Biology,*, vol. 48, no. 3, pp. 443–453, 1970.

[28] J. West, A. Haake, E. Rozanski, and K. Karn, ""eyepatterns: Software for identifying patterns and similarities across fixation sequences," *Eye Tracking Research and Applications (ETRA) Symposium. ACM, San Diego, CA*, pp. 149–154, 2006.

[29] T. Smith and M. S. Waterman, "Identification of common molecular subsequences," *Journal of Molecular Biology*, vol. 147, no. 1, pp. 195–197, 1981.

[30] L. Stark and S. Ellis, "Scanpaths revisited: Cognitive models direct active looking," *Eye Movements: Cognition and Visual Perception*, 1981.

[31] S. Ellis and J. Smith, "Pattern of statistical dependency in visual scanning," *Eye Movements and Human Information Processing*, 1985.

[32] D. Noton and L. W. Stark, "Scanpaths in saccadic eye movements while viewing and recognizing patterns," *Vision Research*, vol. 1, pp. 929–942, 1971.

[33] S. R. Ellis and L. Stark, "Statistical dependency in visual scanning," *Human Factors*, vol. 28, no. 4, pp. 421–438, 1986.

[34] S. S. Hacisalihzade, L. W. Stark, and J. S. Allen, "Visual perception and sequences of eye movement fixations: A stochastic modelling approach," *IEEE Trans. Syst., Man, Cybern.*, vol. 22, pp. 474–481, 1992.

[35] R. Pieters, E. Rosbergen, and M. Wedel, "Visual attention to repeated print advertising: A test of scanpath theory," *Journal of Marketing Research, JSTOR*, 1999.