

Review of algorithms for detection of fixations from eye tracker database

Pramodini A.Punde

Assistant Professor,

MGM Dr.G.Y.P. College of C.S. & I.T.

Aurangabad, Maharashtra, India

Dr.Ramesh R. Manza

Associate Professor, Dept. of C.S.and I.T.

Dr.B.A.M.University,

Aurangabad, India

Abstract— Eye movements are a continuous part of sensory perception of a scene. Whenever we interact with the visual environment we generate saccadic eye movements. Saccades move the eyes in a ballistic manner from one point to another, interspersed by fixations where the eye is stable. Scan paths describe the sequence of fixations and saccades. This eye movement signature data can be collected as a part of a person's daily activities and can be analyzed to detect eye diseases, psychological disorders. Eye tracking systems have various applications including neurological, psychological, usability studies, biomedical applications, and Human Computer Interaction applications. In this paper we discuss various algorithms to compute fixations from eye tracker data.

Keywords—Eye tracking; fixations; gaze position

I. INTRODUCTION

There has been significant advancement in eye tracking technology recently. Current eye trackers have become affordable, accurate and easy to use. These eye trackers do not require the head to be constrained in a specific position and can collect the data in standard viewing postures. This has enabled large scale low-cost availability of such data from multiple subjects especially in multimedia settings. This information can be extremely beneficial to computer vision algorithms and eye tracking guided computer vision algorithms have become a relevant topic of interest. These hybrid techniques have two design challenges. The first challenge is to effectively identify useful information from the eye tracking data for the task (computer vision problem) one is trying to solve. Secondly, a guiding mechanism needs to be designed that utilizes this eye tracking information to further aid the computer vision task at hand.

Eye movements are a continuous part of sensory perception of a scene. Whenever we interact with the visual environment we generate saccadic eye movements. Saccades move the eyes in a ballistic manner from one point to another, interspersed by fixations where the eye is stable. Scan paths describe the sequence of fixations and saccades. This eye movement signature data can be collected as a part of a person's daily activities and can be analyzed to detect eye diseases, psychological disorders. Eye tracking systems have various applications including neurological, psychological, usability studies, biomedical applications, Human Computer Interaction applications. In this paper we discuss various algorithms to compute fixations from eye tracker data.

The rotation of the eyes is controlled by six muscles, giving it three degrees of freedom in relation to the head. However, Donders' law states that due to neural constraints there is a unique orientation for each gaze direction. Effectively this gives the eye only two degrees of freedom. It is often relevant to talk about the duration and the amplitude of eye movements however, these values can differ significantly between investigations due to different recording methods.

1.1 Saccade

Due to the limited size of the fovea, the eyes frequently move around, placing different areas of interest in the fovea, during which the brain progressively builds a mental image of the visual field. The most common way of repositioning the gaze position is by a rapid sweep called a saccade. The saccade is a ballistic movement, meaning it is pre-programmed and does not change once it has started. Saccades of amplitude 40° peak at velocities of 300–600°/s and last for 80–150 ms [5]. During the movement our conscience mind rejects the input data from the eyes, a process called saccadic suppression. A wobbling movement with an amplitude of about 0.15° [6] can sometimes be seen at the end of saccades. This movement is called dynamic overshoot when lasting 10–30 ms and glissadic overshoot when lasting 30–500 ms [7]. They are usually grouped together and referred to as simply glissades. A common belief is that they have no useful function and are simply an

imperfection [6]. The saccade has a refractory period of 200 ms, during which a new saccade cannot be issued [8].

1.2 Fixation

Between saccades the gaze position is kept fairly fixed in what is called a fixation; in reality though, the eyes never stop moving. If they did, our perception of the visual field would rapidly fade and become homogeneous due to neural adaptation. To cope with this the eyes are constantly in motion, but the movements are normally too small and fast to be perceived without a high-precision and high speed eye tracker. The movements that happen during a fixation are:

1.3 Tremor: The smallest known eye movement is an oscillation called ocular micro tremor.

Consensus has not been reached regarding the exact purpose of the tremor. The movement is too small to be measured by most eye trackers.

1.4 Drift: A fixation starts with the object of interest centered in the fovea; it will, however, slowly drift away, in the range 2–15 min (as cited in Martinez et al., 2004[4]). 1 min (arc minute) equals 1/60 degree.

1.5 Micro saccade: A small movement with comparable amplitude to the drift, but with the speed and characteristics of a saccade. Some believe its sole purpose is to re-centre the object of interest in the fovea after drifts, but consensus has not been reached [4].

It is experimentally confirmed that the eyes constantly have to move to maintain a strong neural response, but exactly which of those movements achieve this is highly disputed. Some believe tremors are too small to achieve this, others disagree. Some believe drifts achieve this, others say they are merely byproducts of imperfections in the oculomotor system and fill no useful purpose. Even micro saccades are seen as both realignments of the object to the fovea and a refreshment of stimulus to keep the neural signals from fading [9].

Eye tracking mechanism:

The raw data from the eye tracker consist of gaze points represented as x and y coordinates on the viewing plane; for eye trackers with a built-in screen this is pre-defined as the screen. The classification algorithm tags every point with the appropriate label (e.g. saccade) and an ordinal number. The ordinal number describes to which movement a point belongs: this means the points in the same fixation will have the same number. This method is used because some algorithms make it possible for two consecutive frames that were labeled as fixations to be part of different fixations, in which case they are given different ordinal numbers. However, most algorithms employ the method used in Salvucci and Goldberg (2000) [16] where neighboring points with the same label always belong to the same movement. In this paper we consider the most important element of eye tracking data *fixation*.

II. ALGORITHMS FOR EYE TRACKING DATA

Information hidden in the eye movement signal can be a valuable source of knowledge about a human mind. This information is commonly used in multiple fields of interests like psychology, medicine, business, advertising or even software developing. The proper analysis of the eye movement signal requires its elements to be extracted. The most important ones are fixations - moments when eyes are almost stable and the brain is acquiring information about the scene. There were several developed algorithms for detecting fixations.

A. *Velocity-based algorithms (I-VT)*

Velocity-threshold fixation identification (I-VT) is the simplest of the identification methods to understand and implement. I-VT is a velocity-based method that separates fixation and saccade points based on their point-to-point velocities [e.g., 5, 18]. The velocity profiles of saccadic eye movements show essentially two distributions of velocities: low velocities for fixations (i.e., <100 deg/sec), and high velocities (i.e., >300 deg/sec) for saccades. This aspect of saccadic eye movements makes velocity-based discrimination fairly straightforward and robust. While it is possible to create a locally adaptive velocity-based fixation identification algorithm, velocity profiles have strong physical and physiological underpinnings, and thus static criteria are usually sufficient here. I-VT begins by calculating point-to-point velocities for each point in the protocol. Each velocity is computed as the distance between the current point and the next (or previous) point. I-VT then classifies each point as a fixation or saccade point based on a simple velocity threshold: if the point's velocity is below threshold, it becomes a fixation point, otherwise it becomes a saccade point. The process then collapses consecutive fixation points into fixation groups and discards saccade points. Finally, I-VT translates each fixation group to a representation $\langle x, y, t, d \rangle$ using the centroid (i.e., center of mass) of the points as x and y, the time of the first point as t, and the duration of the points as d.

I-VT requires the specification of one parameter, the velocity threshold. If angular velocities can be computed (i.e., the distance from eye to visual stimuli is known), the point-to-point velocity threshold can be approximated from a reasonable angular velocity threshold [6]; for instance, Sen and Megaw [18] used a threshold of 20 degrees/second. However, when only point-to-point velocities are known, an appropriate value for the velocity threshold may need to be inferred based on aspects of data collection (e.g., sampling frequency) along with some exploratory data analysis.

B. Dispersion-Threshold Identification (I-DT)

In contrast to the velocity-based identification of I-VT and IHMM, dispersion-threshold identification (I-DT) utilizes the fact that fixation points, because of their low velocity, tend to cluster closely together. I-DT identifies fixations as groups of consecutive points within a particular dispersion, or maximum separation [e.g., 19]. Because fixations typically have duration of at least 100 ms, dispersion-based identification techniques often incorporate a minimum duration threshold of 100-200 ms to help making equipment variability easy. The following I-DT algorithm is based on Widdel's data reduction algorithm.

The I-DT algorithm uses a moving window that spans consecutive data points checking for potential fixations. The moving window begins at the start of the protocol and initially spans a minimum number of points, determined by the given duration threshold and sampling frequency. I-DT then checks the dispersion of the points in the window by summing the differences between the points' maximum and minimum x and y values; in other words, dispersion $D = [\max(x) - \min(x)] + [\max(y) - \min(y)]$. Note that alternative dispersion metrics could be based upon spatial variance or area of samples. If the dispersion is above the dispersion threshold, the window does not represent a fixation, and the window moves one point to the right. If the dispersion is below the dispersion threshold, the window represents a fixation. In this case, the window is expanded (to the right) until the window's dispersion is above threshold. The final window is registered as a fixation at the centroid of the window points with the given onset time and duration. This process continues with window moving to the right until the end of the protocol is reached.

This characterization of fixations uses the centroid and diameter. A circular area is usually assumed, and the mean distance from each sample to the fixation centroid provides an estimate of the radius. Also, dispersion-based algorithms are sometimes used to locate clusters within minimum spanning tree network representations. The graph forms an efficient framework for rapid search of large sample sets.

The I-DT algorithm requires two parameters, the dispersion threshold and the duration threshold. Like the velocity threshold for I-VT, the dispersion threshold can be set to include $1/2^\circ$ to 1° of visual angle if the distance from eye to screen is known. Otherwise, the dispersion threshold can be estimated from exploratory analysis of the data. The duration threshold is typically set to a value between 100 and 200 ms, depending on task processing demands. Note that this characterization of I-DT is conceptually different from some dispersion-based methods that center on clustering algorithms. For instance, the leader algorithm and the k-means algorithm are two such methods [8]. However, most of these algorithms require an initial estimate of the number of clusters in the protocol, making them less useful for eye-movement data, where the number of fixations is typically unknown.

C. HMM Identification (I-HMM)

Hidden Markov model fixation identification (I-HMM) uses probabilistic analysis to determine the most likely identifications for a given protocol [15, 16]. Hidden Markov models (HMMs) are probabilistic finite state machines that have been employed extensively in the fields of speech and handwriting recognition. I-HMM uses a two-state HMM in which the states represent the velocity distributions for saccade and fixation points. This probabilistic representation helps I-HMM perform more robust identification than a fixed-threshold method such as I-VT.

Hidden Markov model (HMM) is a statistical model that is used to determine the sequence of states that corresponds best with a sequence of observable data (the signal). The sequence of states the correct classification of the eye movements is unknown, or hidden. The signal the eye tracking data is used to determine the most likely hidden sequence using a probabilistic model. First, observational probabilities describe how likely it is that a signal corresponds to a certain state, much like how threshold-based methods work. The signal is converted to relevant and manageable data called features. If the velocity feature from I-VT, v , is used to classify fixations and saccades, a small v , meaning the eye gaze is slow-moving or stationary, should give a high probability of a fixation and a low probability of a saccade. In addition to that, transitional probabilities describe how likely the transition between two states is. For instance, if classifying fixations, saccades and glissades, the probability of a transition from a fixation to a glissade can be set next to zero, as glissades by definition follow saccades. The probabilities are many (typically $n^2 + 3n$ for an n -state model) and are best determined automatically by training.

The HMM includes two sets of probabilities: observation and transition probabilities. The observation probabilities for each state (the small v distributions) represent the distribution of expected velocities in that state. The first state represents saccade points, and thus contains a distribution centered on higher velocities; the second state represents fixation points, and thus contains a distribution centered on lower velocities. The transition probabilities for each state (the arrows exiting the states) represent the likelihood of remaining in the state or making a transition to another state. Thus, the HMM provides a probabilistic representation of the observations (i.e., velocities) generated during saccadic eye movements. The parameters of I-HMM comprise the observation and transition parameters in the two-state HMM: two observation probability parameters (the mean and variance of the distribution) and two transition probabilities for each state, for a total of eight parameters. While this parameter space is certainly more complex than that of I-VT, the parameters of IHMM can be learned through a procedure called reestimation [13]. Given a training set (i.e., a set of protocols) reestimation learns the values of HMM parameters that maximize the probability of the training set given the HMM. Thus, with protocol data from a particular eye-tracking setup, reestimation can provide parameter values that, once estimated, can be used for any protocols collected in that setup.

D. MST Identification (I-MST)

MST identification (I-MST) is based on minimum spanning trees (MSTs) that is, a tree connecting a set of points such that the total length of the tree's line segments is minimized. MSTs can provide a highly flexible and controllable representation for dispersion-based fixation identification [7]. A two-step approach is necessary, first requiring the construction of the MST followed by a search of the MST. Construction uses Prim's algorithm [2]. There is one and only one MST for a set of points. The advantage of an MST data representation lies in the degree of control, flexibility, and local adaptation for dispersion analysis, as well as improved subsequent characterizations of defined fixations. Fixation identification requires traversing the already defined MST. A depth-first search is made to determine the maximum depth of interconnectivity at each point. Branching depths below a defined set point signal locations near the edge of the MST that are not appropriate candidates for separating fixations. If the edges connected to each endpoint exceed the defined minimum branching depth, then the network of edges connected to each endpoint is gathered into an associated edge length distribution. The mean μ and standard deviation σ of edge lengths provide a locally adaptive comparison for separation of fixations. Separation can occur based upon comparison of the edge under consideration of both μ and σ of neighbor edge lengths. In this way, a framework is achieved for controlling where fixations may occur in an MST and for determining how local adaptivity affects fixation decisions. The MST format allows additional characterization parameters for fixations. For example, the MST length is defined by the longest path through the graph. Critical paths can be defined by areas with minimal branching structure. These and other parameters allow an estimate of the shape and possibly direction of fixations; these go beyond simpler centroid and size characterizations.

E. Area-based Algorithms

Area-of-Interest Identification (I-AOI)

The four previous identification methods can identify fixations at any location in the visual field. In contrast, area-of-interest fixation identification (I-AOI) identifies only fixations that occur within specified target areas. The target areas are rectangular regions of interest that represent units of information in the visual field. These target areas, generally used in later analyses like tracing, keep identified fixations close to relevant targets. I-AOI also utilizes a duration threshold to help distinguish fixations in target areas from passing saccades in those areas. I-AOI begins by associating data points with target areas: it labels points within a target area as a fixation point for that target, and labels points outside of all target areas as saccades. I-AOI then collapses consecutive fixation points for the same target into fixation groups, discarding saccade points. Finally, it removes fixation groups that fall below a given duration threshold and transforms each fixation group into a fixation tuple. As explained earlier, I-AOI can provide a larger picture than fixations alone. Like a fixation, dwell time within an area has a starting time and an ending time. However, since fixations can serve as the input data into AOI determination, the time between these dwells do not describe saccadic behavior. Rather, there could be multiple saccades and fixations interspersed between AOIs. While AOI-based dwell-time algorithms are not fixation algorithms per se, they are valuable as a concept to help explain higher-level collections of fixations organized about visual targets and areas.

III. EXPERIMENTAL WORK:

Fixation generation

In the present paper, we used DOVES database for our study. Sampled voltages corresponding to the eye movements of the observers for each trial were converted to gaze coordinates (i.e. position of gaze on the image

in pixels). Next, the path of the subject's gaze was divided into fixations and the intervening saccadic eye movements using spatio-temporal criteria derived from the known dynamic properties of human saccadic eye movements. Stated simply, a sequence of eye position recordings was considered to constitute a fixation if the recorded gaze coordinates remained within a stimulus diameter of 1° visual angle for at least 100 ms. The exact algorithm (adapted from ASL, 1998) accounts for drifts, blinks and micro-saccades. A mean of 12.1 (SD = 3.0) fixations per image was observed. Patterns of fixations for four single trials for one observer are shown in Fig. 1. The serpentine lines show the eye movement trajectories linking serial fixations (marked by circles). The first (forced) fixation is denoted by a star near the centre of the image. Since there are several techniques to compute fixations from eye movement traces, the online database also provides users with the raw eye movement data sampled at 200 Hz allowing the application of other fixation detection algorithms.

Eye movement data may be subjected to a wide variety of statistics to determine saccade lengths, angles, speeds and fixation locations, durations and degrees of clustering, to name but a few. Whilst individual experiments may ascribe specific meanings to particular eye movement statistics, their general distributions are of great importance to applications such as accurate modeling/prediction of human gaze patterns, and in experimental procedures for which deviations from expected values in any of these measurements are to be interpreted as indicative of the observer's cognitive state.

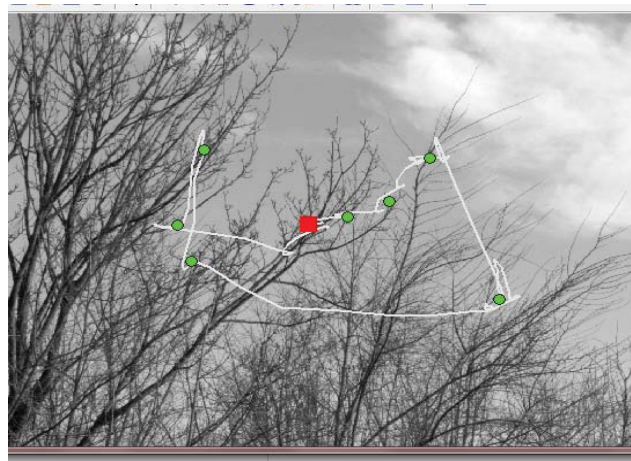


Fig.1. Fixations of first subject. Red square shows first fixation. Green dots show computed fixations. White line shows the "raw" eye movement trace.

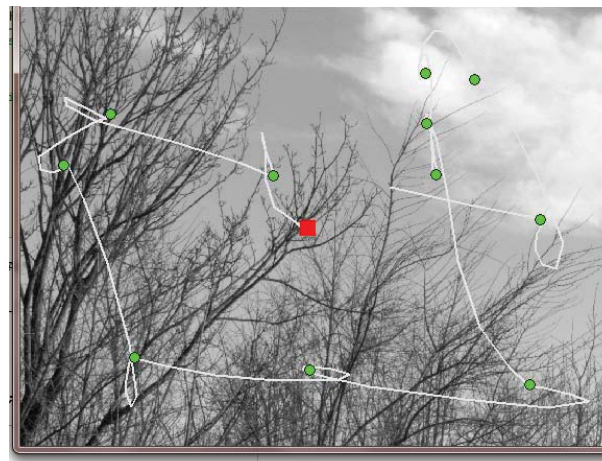


Fig.2. Fixations of second subject.

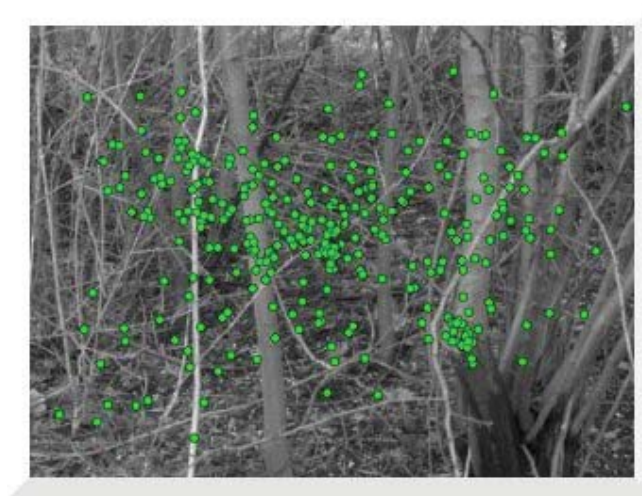


Fig. 3. An example of fixations from all subjects on an image

IV. CONCLUSION

It is well known that eye movements are also an integral part of the encoding of visual stimuli; human eyes actively interact with their visual environment, gathering information from the foveated (variable spatial resolution) visual input using a combination of steady eye fixations linked by rapid ballistic eye movements called saccades.

In the present paper, various algorithms for computing fixations from eye tracker data have been studied namely, velocity based (I-VT), Dispersion based (I-DT), Hidden-Markov Model, and area based algorithm. We used DOVES database to compute fixations.

Recent studies have greatly improved our understanding of how eye movements are deployed in real-world scene viewing, during reading and information processing, visual search, and during natural tasks requiring coordinated eye and body movements. Indeed, the study of eye movements is a bright research area with a variety of applications ranging from cognitive psychology to computation, neuroscience and business. Most of the researchers have used eye movement metrics for viewing eye patterns for abnormalities and psychological disorders and some researchers used it for machine learning, usability research and HCI.

REFERENCES

- [1] van der Linde, I., Rajashekar, U., Bovik, A.C., Cormack, L.K. (2009). DOVES: A database of visual eye movements. *Spatial Vision*, 22(2): 161-177. URL: <http://live.ece.utexas.edu/research/doves>
- [2] S. J. Landry, T. B. Sheridan, and Y. M. Yufik, "A methodology for studying cognitive groupings in a target tracking task," *IEEE Trans. Intell. Transp. Syst.*, vol. 2, no. 2, pp. 92-100, Jun. 2001
- [3] P. Chapman, G. Underwood, and K. Roberts, "Visual search patterns in trained and untrained novice drivers," *Transp. Res. Part F, Psychol. Behav.*, vol. 5, no. 2, pp. 157-167, 2002.
- [4] G. Underwood, P. Chapman, N. Brocklehurst, J. Underwood, and D. Crundall, "Visual attention while driving: Sequences of eye fixations made by experienced and novice drivers," *Ergonomics*, vol. 46, no. 6, pp. 629-646, 2003.
- [5] A. C. Oltekin, S. I. Fabrikant, and M. Lacayo, "Exploring the efficiency of users' visual analytics strategies based on sequence analysis of eye movement recordings," *Int. J. Geog. Inf. Sci.*, vol. 24, no. 10, pp. 1559-1575, 2010.
- [6] J. Tchalenko, "Segmentation and accuracy in copying and drawing: Experts and beginners," *Vis. Res.*, vol. 49, no. 8, pp. 791-800, 2009.
- [7] W. J. Gibbs and R. S. Bernas, "Visual attention in newspaper versus TV oriented news websites," *J. Usability Stud.*, vol. 4, no. 4, pp. 147-165, 2009.
- [8] C. Holland and O. Komogortsev, "Biometric identification via eye movement scanpaths in reading," in *Proc. Int. Joint Conf. Biometrics*, 2011, pp. 1-8.
- [9] C. Krischer and W. Zangemeister, "Scanpaths in reading and picture viewing: Computer-assisted optimization of display conditions," *Comput. Biol. Med.*, vol. 37, no. 7, pp. 947-956, 2007.
- [10] A. Poole, L. J. Ball, and P. Phillips, "In search of salience: A resposetime and eye-movement analysis of bookmark recognition," in *People and Computers XVIII-Design for Life*, S. Fincher, P. Markopoulos, D. Moore, and R. Ruddle, Eds. London, U.K.: Springer, 2005, pp. 363-378.
- [11] R. Caldara and S. Miellet, "iMap: A novel method for statistical fixation mapping of eye movement data," *Behav. Res. Methods*, vol. 43, no. 3, pp. 64-878, 2011.
- [12] R.W. Reeder, P. L. Pirulli, and S. K. Card, "WebEyeMapper and WebLogger: Tools for analyzing eye tracking data collected in web-use studies," in *Proc. ACM Conf. Human Factors Comput. Syst.*, New York, NY, USA, 2001, pp. 19-20.
- [13] F. Cristino, S. Mathot, J. Theeuwes, and I. D. Gilchrist, "ScanMatch: A novel method for comparing fixation sequences," *Behav. Res. Methods*, vol. 42, no. 3, pp. 692-700, 2010.

- [14] J. Goldberg and X. Kotval, "Computer interface evaluation using eye movements: Methods and constructs," *Int. J. Ind. Ergonom.*, vol. 24, no. 6, pp. 631–645, 1999.
- [15] D. Noton and L. Stark, "Eye movements and visual perception," *Sci. Amer.*, vol. 224, no. 6, pp. 34–43, 1971.
- [16] S. Brandt and L. Stark, "Spontaneous eye movements during visual imagery reflect the content of the visual scene," *J. Cognit. Neurosci.*, vol. 9, no. 1, pp. 27–38, 1997.
- [17] A. T. Duchowski, J. Driver, S. Jolaoso, W. Tan, B. N. Ramey, and A. Robbins, "Scanpath comparison revisited," in *Proc. Symp. Eye-Tracking Res. Appl.*, New York, NY, USA, 2010, pp. 219–226.
- [18] C. M. Privitera and L. W. Stark, "Algorithms for defining visual regions of -interest: Comparison with eye fixations," *IEEE Trans. Pattern Anal.Mach. Intell.*, vol. 22, no. 9, pp. 970–982, Sep. 2000.
- [19] P. Hejmady and N. H. Narayanan, "Visual attention patterns during program debugging with an IDE," in *Proc. Symp. Eye Tracking Res. Appl.*, 2012, pp. 197–200.