Real-time facial expression recognition to track non-verbal behaviors as lie indicators during interview

Arif Budi Setiawan [a,1,*], Kaspul Anwar [a,2], Laelatul Azizah [b,3], Adhi Prahara [c,4]

[a] Electrical Engineering Department, Universitas Ahmad Dahlan, Jl. Ring Road Selatan, Tamanan, Banguntapan, Bantul, Yogyakarta 55191, Indonesia

[b] Department of Psychology, Universitas Ahmad Dahlan, Semaki, Umbulharjo, Kota Yogyakarta, Yogyakarta 55191, Indonesia

[c] Informatics Department, Universitas Ahmad Dahlan, Jl. Ring Road Selatan, Tamanan, Banguntapan, Bantul, Yogyakarta 55191, Indonesia

[1] absetiawan00@gmail.com; [2] kaspulrisma@gmail.com, [3] azkiatunn@gmail.com, [4] adhi.prahara@tif.uad.ac.id*

* corresponding author

ARTICLE INFO

ABSTRACT

During interview, a psychologist should pay attention to every gesture and response, both verbal and nonverbal language/behaviors, made by the client. Psychologist certainly has limitation in recognizing every gesture and response that indicates a lie, especially in interpreting nonverbal behaviors that usually occurs in a short time. In this research, a real time facial expression recognition is proposed to track nonverbal behaviors to help psychologist keep informed about the change of facial expression that indicate a lie. The method tracks eye gaze, wrinkles on the forehead, and false smile using combination of face detection and facial landmark recognition to find the facial features and image processing method to track the nonverbal behaviors in facial features. Every nonverbal behavior is recorded and logged according to the video timeline to assist the psychologist analyze the behavior of the client. The result of tracking nonverbal behaviors of face is accurate and expected to be useful assistant for the psychologists.

1. Introduction

Interview or counseling is one of the important tools to support the decision of psychologist. During interview or counseling, psychologists need to pay attention to the verbal and nonverbal language/behaviors of the client (the person being interviewed), especially the client's nonverbal behaviors. However, the ability of psychologists to observe the behavior is greatly limited by the lack of [1]:

a. **Sensitivity and exercise**. Lack of experience makes the psychologists have low sensitivity. Therefore, during an interview or counseling session, a lot of nonverbal information is missed from observation.

b. **The ability to capture the meaning of the client's gesture expression**. Psychologists often experience human error. This makes the psychologists did not recognize the nonverbal language of the client.

c. **Psychologists' knowledge in interpreting the client's nonverbal behaviors during interview or counseling**. Limitation of knowledge related to nonverbal behaviors that possessed by the psychologists can hinder the process of interpretation. This makes the goal is not achieved properly.

Nonverbal behavior does not appear randomly. The client continues to present the behavior along with the verbal language. Psychologists should record every implied nonverbal behavior during

interview or counseling process to identify whether the client is telling the truth or lies. Therefore, the idea to assist psychologists' decision during interview are proposed by researchers [2, 3].

Polygraph [2], a lie detector that measures and records blood pressure, pulse, respiration, and skin conductivity while a person is asked and answers a series of questions. Polygraph is used as an interrogation tool that famous agencies such as FBI, NSA and CIA are reported to use Polygraph to interrogate suspects and screen their new employees.

Mottelson *et al.* [3] introduce Veritaps: a communication device to help users identify the truths and lies. The use of features such as touch pressure and the inter-tap details of number entry shows the potential of using mobile interaction as a truth estimator. The result is 0.98 of F1-score for classifying truth and 0.57 for lies.

For a specific nonverbal behavior of face, facial expression plays important role. This is supported by the significant development of face expression recognition method. Sarode and Bhatia [4] implement 2D appearance-based local approach for the extraction of intransient facial features and recognition of four facial expressions. Radial Symmetry Transform and edge projection analysis use for feature extraction then create a dynamic spatio-temporal representation of the face. The classification process categorized into one of the expression classes. The result is 81% accuracy for facial expression recognition from grayscale image.

Dhavalikar and Kulkarni [5] propose an Automatic Facial Expression Recognition System (AFERS). The proposed method has three stages namely face detection, feature extraction and facial expression recognition. The facial features like eyes, nose, and mouth are extracted using AAM (Active Appearance Model) method. They achieve 90%-95% recognition accuracy using minimum Euclidean distance between training images and test image then after modification using ANFIS (Artificial Neuro-Fuzzy Inference System) the accuracy raised to 100%.

Lopes *et al.* [6] propose a combination of Convolutional Neural Networks and image processing to recognize facial expression. Due to the lack of available dataset for facial expression recognition with deep architectures, they apply pre-processing techniques such as rotation correction, cropping, down-sampling and intensity normalization to extract only expression specific features from a face image and explore the presentation order of the samples during training. They achieve 96.76% accuracy in the CK+ dataset and claim that the proposed method is fast to train and allows real time facial expression recognition.

This research proposes a real time facial expression recognition to track nonverbal behaviors that indicate a lie during interview or counseling. The proposed method uses face detection to find the client's face, face landmark recognition to extract facial features, and specific image processing steps to extract nonverbal behaviors that indicate a lie from each facial features. The rest of this paper is organized as follow: Section 2 presents the facial expression recognition algorithm. Section 3 presents the procedure of the proposed method. Section 4 presents the results and discussion. Finally, the conclusion of this work is described in Section 5.

2. The Facial Expression Recognition

2.1. Nonverbals of the face

In term of psychology, a lie is a statement used intentionally for the purpose of deception. Lies can be identified from verbal or nonverbal language. Nonverbal language/behavior is unconscious motion of the body such as gestures, facial expression, tone/vibration of voice, breath, and so on. Verbal language can be contrary to nonverbal language and nonverbal language can support or emphasize verbal language [1].

In nonverbal behaviors, there are some facial expressions that indicate a lie. Among them are false smile, diminished pupils, avoiding eye contact, high frequency of winks, furrow in the forehead and nose, and raised eyebrows [7]. Table 1 shows some of the nonverbal behaviors of face that indicate a lie. Fig. 1 shows the facial expression of nonverbal behaviors that indicate a lie.

Table 1. The nonverbal behaviors of face that indicate a lie.

No	Non-verbal behavior	Description
1.	Avoiding eye contact	Eye sight that always avoid the eyes of the other person. This is a response from two limbic brain that indicates an escape attempt from something that is considered threatening.
2.	Diminished pupil	Diminished pupils indicate a stress on a person.
3.	Eye blocking	Eye blocking is also a limbic brain response which is an effort to avoid a threat or discomfort feeling.
4.	Furrow in the forehead or nose	Furrow in the forehead or nose is an indication of high stress and it is believed to be a gesture for lying.
5.	False smile	False smile is a fake smile. It indicates by attracted muscle in the corner of the mouth towards horizontal or the direction of the ear.
6.	Raised eyebrows	Raised eyebrows is an indication of fear.

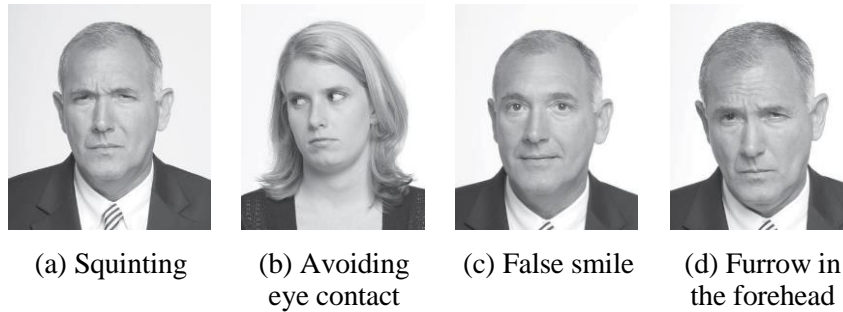| (a) Squinting | (b) Avoiding eye contact | (c) False smile | (d) Furrow in the forehead |

Fig. 1. Facial expressions that indicate negative emotions or a lie [7].

2.2. Face detection

One of the successful algorithm for face detection is Haar-like feature-based cascade classifiers [8]. Haar-like features are extracted using convolutional kernel as shown in Fig. 2 where each feature is a single value obtained by subtracting sum of pixels under the white rectangle from sum of pixels under the black rectangle. The sum calculation can be done using integral image.

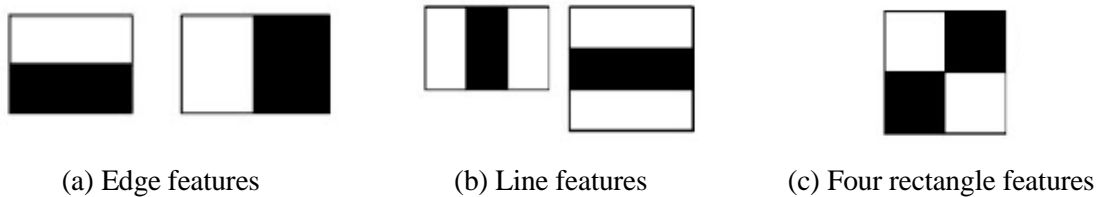| (a) Edge features | (b) Line features | (c) Four rectangle features |

Fig. 2. Haar-like features convolutional kernel [8].

Haar-like features extraction resulted in large features. Adaboost is used to reduce and select only relevant features. The method applies each feature on the training images. For each feature, finds the best threshold which will classify the faces to positive and negative. The features with minimum error rate are selected. The final classifier is a weighted sum of some weak classifiers.

To improve the method, [8] uses cascade of classifiers. Instead of applying all the features on a window, it groups the features into different stages of classifiers and apply one-by-one. If a window fails in the first stage, then discard the window. If a window passes the first stage, apply the second stage of features and continue the process. The window which passes all stages is a face region.

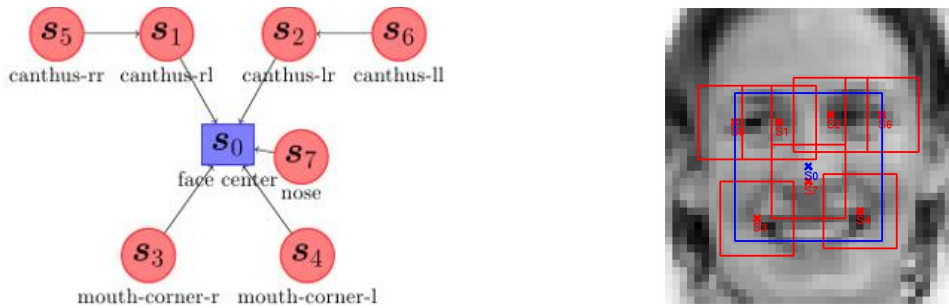## 2.3. Facial Features Extraction

Facial features extraction uses facial landmark recognized from a structured output classifier based on the Deformable Part Models (DPM). The quality of landmark configuration $s = (s_0, \ldots, s_{M-1}) \in S$ for given image I is measured by scoring function $f: I \times S \to R$. The scoring function is defined as

a sum of the appearance fit and the deformation cost. Exact formulation of f is derived from the graph constraints as shown in equation (1) [9].

$$f(I, s) = \sum_{i=0}^{M-1} q_i(I, s_i) + \sum_{i=1}^{M-3} g_i(s_0, s_i) + g_5(s_1, s_5) + g_6(s_2, s_6) + g_7(s_0, s_7) \qquad (1)$$

Functions $q_i(I, s_i)$ which is the appearance fit and $g_i(s_i, s_j)$ which is the deformation cost are parameterized. These parameters are learned from annotated examples using the structured output Support Vector Machine (SVM) algorithm. Fig. 3 shows the facial landmark detection using 8 components. From Fig. 3, (a) is the graph constraints model with 8 components and (b) is the facial landmark detection on face. Fig. 4 shows the block diagram of facial landmark detection.

The facial features are extracted using facial landmark information. Before that, we divide the face region into three part namely upper, middle, and lower face. Forehead is extracted from the upper face area which is 1/3 of face area. Eyes area is extracted from the corner eyes landmark and belong to the middle face area. Mouth also extracted from the mouth corner landmark and belong to the lower face area.

(a) Graph constraints model of facial landmark.	(b) Facial landmark detection on face.

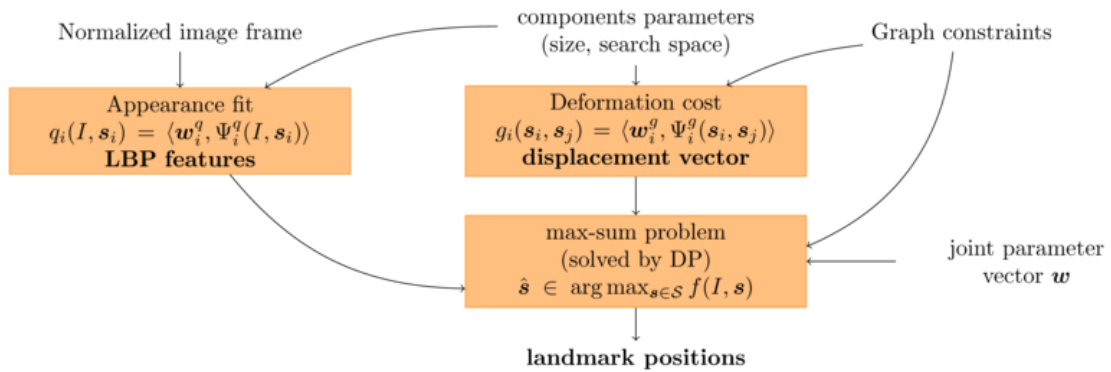Fig. 3. Facial landmark detection using 8 components [9].

Fig. 4. Block diagram of facial landmark detection [9].

2.4. Facial Expression Recognition

The nonverbal behavior is tracked from the facial expression. As described in Table 1, we choose three of the nonverbal behaviors that indicate a lie namely avoiding eye contact, furrow in the forehead, and false smile.

a. **Avoiding eye contact.** Eye sight is tracked to identify avoid eye contact behavior using movement of pupils. Template matching method is used to detect pupils in the eyes area that estimated from facial landmark. Template matching method requires a source image and template image that will be compared to detect the highest matching area using sliding windows technique and correlation coefficient method as shown in equation (2). The pupil area indicated from window that has the highest matching score.

$$R(x, y) = \sum_{x', y'} \left(T'(x', y') \cdot I(x + x', y + y') \right) \qquad (2)$$

Where $R(x, y)$ is the score of coefficient correlation in (x, y), T is the template image and I is the source image.

b. **Furrow in the forehead**. Furrow in the forehead can be detected using line detector algorithm. We use Sobel horizontal kernel as shown in equation (3) to extract the horizontal orientation of gradient using convolution and approximate the line using Hough transform.

$$G_x = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} \tag{3}$$

Hough transform expresses lines in the Polar coordinate. For each point (x_0, y_0), the family of lines that goes through that point can be expressed as equation (4). This means each pair (r_θ, θ), represents each line that passes by (x_0, y_0) and the plot of family of lines that goes through that point is in the form of sinusoid curve.

$$r_\theta = x_0 \cos\theta + y_0 \sin\theta \tag{4}$$

If all the points in an image is calculated and if the curves of two different points intersect in the plane $\theta - r$, then that both points belong to the same line. Therefore, a line can be detected by finding the number of intersections between curves. The number of intersection can be used as a threshold to detect line. More efficient version of this algorithm is described in [10] where the method minimizes the amount of computation needed to detect lines by exploiting the difference in the fraction of votes needed to reliably detect lines with different numbers of supporting points.

c. **False smile.** False smile is indicated by the corner of the mouth that stretch towards horizontal or the direction of the ear. False smile can be detected by calculating the difference between the length of normal mouth and the current mouth in horizontal direction. If the difference is greater than the threshold, then it will be considered as false smile. The length of the mouth can be measured using mouth corner detection from mouth landmark recognition in Fig. 3 (b).

3. Method

The procedure of real time facial expression recognition to track nonverbal behaviors that indicate a lie during interview or counseling is shown in Fig. 5. The method takes input from camera stream or video. Face detection method is applied for each frame grabbed from the input. For each face found in the frame, facial landmark recognition is used to extract facial features. We focus on three facial features namely forehead, eyes, and mouth to track the nonverbal behavior such as furrow in the forehead, avoiding eyes contact and false smile.
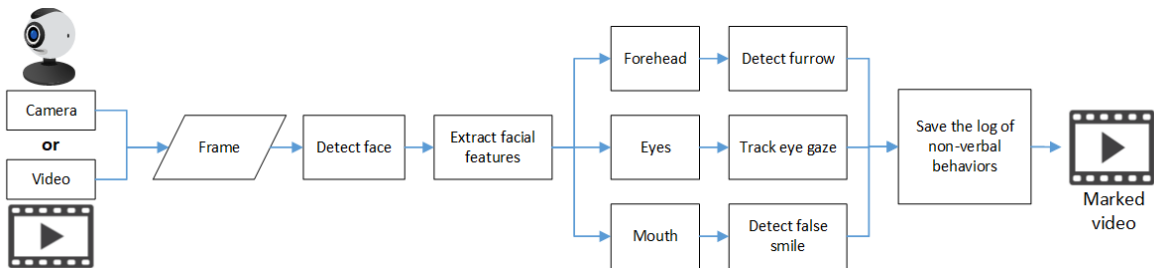

Fig. 5. General procedure of the proposed method.

To assist the psychologists in identifying the nonverbal behaviors of the client, every tracked nonverbal behavior is saved into a log of timeline and status. The psychologist can review the recorded video while paying attention to the video and read the log when the nonverbal behaviors occurs and check the status of nonverbal behaviors such as "furrow in the forehead", "left eye and/or right eye is avoiding contact", and "false smile".

4. Results and Discussion

The proposed method is built using C# with additional image processing library and runs on laptop with processor Intel Core i3 2.0 GHz, 4 GB of RAM, and graphics card NVIDIA GeForce 940MX. The test data consist of 3 interview videos with resolution 1920x1080 and 5 to 10 minutes duration

that follow specific scenarios. The scenario is preserved as a ground truth of nonverbal behaviors that present during interview.

4.1. Nonverbal Behaviors Detection

The method tracked three nonverbal behaviors namely furrow in the forehead, avoiding eye contact, and false smile. The result of facial features extraction is shown in Fig. 6. From Fig. 6, blue rectangle is the face area, magenta rectangles are the eyes area, cyan rectangle is mouth area, cyan line strip is the forehead area, and the eight yellow dots are facial landmark such as corner of the eye and mouth. As seen in Fig. 6, the facial features extraction method can accurately extract the face, eyes, mouth, and forehead.

Fig. 7 - 9 shows the nonverbal behaviors tracked by the method. Fig. 7 shows the comparison result between (a) normal forehead and (b) furrow in the forehead (marked by cyan horizontal line). From Fig. 7 (b), furrow in the forehead is detected. We apply a threshold > 3 which is the number of wrinkles need to be present in the forehead before it detected as lie indicator and only detect the horizontal line using Hough transform with $\theta = 90°$ from Sobel horizontal oriented gradients.
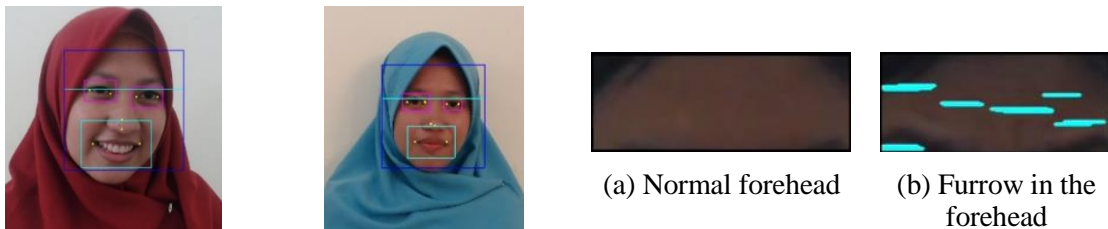

Fig. 6. The result of facial features extraction.

(a) Normal forehead   (b) Furrow in the forehead

Fig. 7. Comparison result between normal and furrow in the forehead behavior.

Fig. 8 shows the comparison result between (a) normal eye sight (marked by green line) and (b) avoiding eye contact behavior (marked by red line). From Fig. 8 (b), avoiding eye contact behavior is detected. The pupils are moved to the left from the eyes center. If the pupils are in the corner of the eyes, then the client is looking away which means avoiding eye contact behavior is occurred.

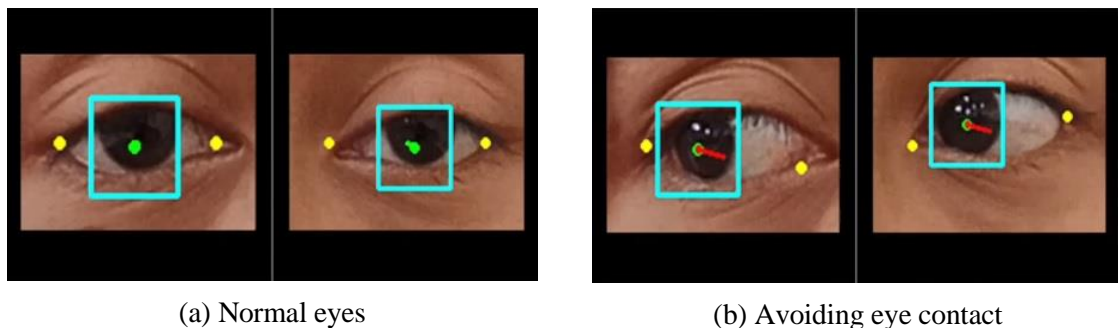(a) Normal eyes   (b) Avoiding eye contact

Fig. 8. Comparison result between normal eye sight and avoiding eye contact behavior.

Fig. 9 shows the comparison result between (a) normal mouth (marked by green line) and (b) false smile (marked by red line). From Fig. 9 (b) false smile is detected because the length of the mouth is changed towards horizontal. The vertical change is ignored because when the client talks, the mouth tends to change towards vertical direction. The length of normal mouth is acquired when the face is present for the first time. If the difference length is greater than threshold of 0.2 * length of normal mouth, then it is considered as false smile behavior.

(a) Normal mouth   (b) False smile

Fig. 9. Comparison result between normal mouth and false smile.

4.2. Performance Evaluation

The proposed method can perform real time. From 5 test videos with resolution 1920x1080, the frame per second is approximately 20. The log of detected nonverbal behaviors is saved with format: [video timeline, status of nonverbal behavior] along with the recorded interview video. The method is fast because it utilizes light algorithms. Only face detection that runs on the whole frame. The other method runs on localized area such as facial landmark recognition which only operates on face area and pupil detection using template matching which only operates on small eyes area.

5. Conclusion

This research proposes real time face expression recognition to track nonverbal behaviors that indicate a lie during interview. The result shows good accuracy when detecting nonverbal behaviors. With good accuracy, this program is expected to assist psychologists in analyzing nonverbal behaviors of the client during interview or during the review of recorded interview video with the log of nonverbal behaviors. For future implementation, we will add another important nonverbal behavior that indicates a lie such as high frequency of blink, eye blocking, and diminishing pupils.

References

[1] S. S. Willis, "Konseling individual teori dan praktek," *Bandung Alf.*, vol. 79, 2004.

[2] J. P. Rosenfeld, "Alternative views of Bashore and Rapp's (1993) alternatives to traditional polygraphy: A critique.," 1995.

[3] A. Mottelson, J. Knibbe, and K. Hornbæk, "Veritaps: Truth Estimation from Mobile Interaction," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018, p. 561.

[4] S. Neeta and S. Bhatia, "Facial Expression Recognition," *Int. J. Comput. Sci. Eng.*, vol. 2, 2010.

[5] A. S. Dhavalikar and R. K. Kulkarni, "Face detection and facial expression recognition system," in *2014 International Conference on Electronics and Communication Systems (ICECS)*, 2014, pp. 1–7.

[6] A. T. Lopes, E. de Aguiar, A. F. De Souza, and T. Oliveira-Santos, "Facial expression recognition with Convolutional Neural Networks: Coping with few data and the training sample order," *Pattern Recognit.*, vol. 61, pp. 610–628, Jan. 2017.

[7] J. Navarro and M. Karlins, *What every body is saying*. HarperCollins Publishers, 2008.

[8] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1, p. I-511-I-518.

[9] M. Uričář, V. Franc, and V. Hlaváč, "Detector of Facial Landmarks Learned by the Structured Output SVM," in *VISAPP '12: Proceedings of the 7th International Conference on Computer Vision Theory and Applications*, 2012, vol. 1, pp. 547–556.

[10] J. Matas, C. Galambos, and J. Kittler, "Robust Detection of Lines Using the Progressive Probabilistic Hough Transform," *Comput. Vis. Image Underst.*, vol. 78, no. 1, pp. 119–137, Apr. 2000.