REAL-TIME EYE RATIO TRACKING FOR DROWSINESS DETECTION BASED ON MACHINE VISION AND FACIAL LANDMARKS

THEO DÕI TỶ LỆ MẮT THEO THỜI GIAN THỰC ĐỂ PHÁT HIỆN NGỦ GẬT SỬ DỤNG CÔNG NGHỆ THỊ GIÁC MÁY DỰA TRÊN ĐIỂM MỐC KHUÔN MẶT

Tran Ngoc Tien[1,*], Tran Van Nghia[1], Pham Thanh Trung[2], Nhu Quy Tho[1], Nguyen Duc Minh[1]

DOI: https://doi.org/10.57001/huih5804.66

ABSTRACT

Bài báo đề xuất thuật toán thời gian thực phát hiện hiện tượng ngủ gật dựa trên camera tiêu chuẩn. Các điểm mốc trên mặt được xác định, đặc biệt là tại vị trí xung quanh vùng mắt của đối tượng. Trên cơ sở đó, một hàm tỷ lệ về khoảng cách các điểm mốc vô hướng được tính toán đại diện cho thời gian mở và nhắm mắt theo thời gian thực. Ngoài ra, bài báo cũng tiến hành khảo sát độ tin cậy của giải pháp được sử dụng trong 3 trường hợp: cường độ ánh sáng thay đổi, đối tượng sử dụng kính, góc nghiêng của khuôn mặt thay đổi so với hướng quét của camera. Kết quả của bài báo có thể được ứng dụng trên các phương tiện giao thông để cảnh báo trạng thái ngủ gật của người điều khiển giúp điều khiển phương tiện an toàn hơn.

Keywords: *Drowsiness detection, facial landmark, Real-time, machine vision.*

TÓM TẮT

The article proposes a real-time algorithm to detect the drowsiness phenomenon based on a standard camera. Face landmarks are identified, especially around the subject's eyes. On that basis, a scale function from extracting scalar quantities of face landmarks is calculated representing real-time eye-opening and closing times. In addition, the article also investigated the reliability of the solution used in three cases: changing light intensity, subject using glasses, face tilt angle changes relative to the scanning direction of the camera. The results of the article can be applied to vehicles to warn drivers of falling asleep to help drive safer vehicle.

Từ khóa: *Phát hiện ngủ gật, điểm mốc khuôn mặt, thời gian thực, thị giác máy.*

[1]Hanoi University of Industry
[2]Vinh Phuc Technical and Economic College
[*]Email: tientn@haui.edu.vn
Received: 20/01/2022
Revised: 15/6/2022
Accepted: 22/11/2022

1. INTRODUCTION

Currently, machine vision technology is widely applied in many fields of the 4th industrial revolution. In which, the problem of face recognition and analysis is the foundation of machine vision technology. The face recognition is developed according to two main trends. Firstly, the studies focus on detecting human faces in images [1-4]. The face detector returns a rectangular bounding box containing a face. The second research direction is to detect and

locate almost exactly facial parts such as eyebrow corners, mount corners, eye corners, etc. The identification of facial landmarks has meaning for the intelligent editing of face images. In addition, it is also used for face recognition, gesture tracking, expression analysis. Tereza Soukupova et al. [5] has been conducted to develop a real-time machine vision algorithm to detect the blinking gesture of the subject. Guarin et al. [6] described an automated face marker according to facial changes in individuals with facial paralysis. Krystyna Malik et al. [7] compared closed and open eyes using the characteristic features of the LBP histogram. In this study, we performed sleep detection. The first step is to detect features on the subject's face using AAMs model [8] and AOM model [9] combined with the available annotations from Multi-PIE Face Database [10]. Then, a mathematical function is built on the basis of the ratio of the distances between the landmarks around the subject's eye area. We evaluate this ratio to determine the time threshold for eyes closed and eyes open. From there, the subject is said to be sleepy if their eyes are closed until a set time. These issues will be discussed in sections 2 and 3 of the paper.

2. FACE CHARACTERISTICS

The objects in an independent AAMs model include the shape and the appearance. Mathematically, the shape is defined as a triangular mesh whose vertices are the coordinates of the feature points on the face given by:

$$\mathbf{s} = \left(x_1, y_1, x_2, y_2, \ldots, x_n, y_n\right)^T \quad (1)$$

where x_i, y_i (i = 1...n)) are the coordinates of the vertices, n is the number of vertices in the mesh and also the corresponding fiducial points are provided for the image.

The shape of the face is transformed by a linear combination of image transformation vectors. The shape **s** expressed as a base shape $\mathbf{s}_0$ plus a linear combination of k eigen-shapes $\mathbf{s}_i$:

$$\mathbf{s} = \mathbf{s}_0 + \mathbf{\Phi}_s \mathbf{c} \quad (2)$$

where $\mathbf{\Phi}_s = (\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_k)$ and the coefficients $\mathbf{c} = (c_1, c_2, ..., c_k)$ are the eigen-shapes and the shape parameters, respectively.

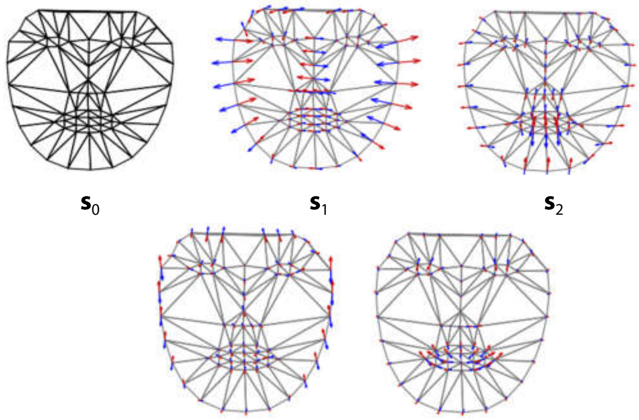
The appearance of the indepent model AAMs is defined within the base mesh $\mathbf{s}_0$. The appearance model is learned by warping of the training images. An appearance **z(x)** can be represented as a base appearance $\mathbf{z}_0(\mathbf{x})$ plus a linear combination of v appearance images $\mathbf{\Phi}_z(\mathbf{x})$:

$$\mathbf{z}(\mathbf{x}) = \mathbf{z}_0(\mathbf{x}) + \mathbf{\Phi}_z(\mathbf{x})\boldsymbol{\lambda} \quad (3)$$

where $\mathbf{\Phi}_z(\mathbf{x}) = (\mathbf{z}_1(\mathbf{x}), \mathbf{z}_2(\mathbf{x}), ..., \mathbf{z}_v(\mathbf{x}))$, the coefficients $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, ..., \lambda_v)$ are the the appearance parameters. Figure 1 shows the shape and appearance model with four face states.

The values $\mathbf{s}_0$ and **s** determine a warp from $\mathbf{s}_0$ to **s**. The warp denote **W(x;c)**. Set **x** is a pixel in $\mathbf{s}_0$, when pixel in the image **I** is **W(x;c)**. The image has the intensity **I(W(x;c))**. The AAMs model estimates the optimal values $\tilde{\mathbf{c}}$ and $\tilde{\boldsymbol{\lambda}}$:

$$\{\tilde{\mathbf{c}}, \tilde{\boldsymbol{\lambda}}\} = \underset{\{\mathbf{c},\boldsymbol{\lambda}\}}{\operatorname{argmin}} \left\| \mathbf{I}(\mathbf{W}(\mathbf{x};\mathbf{c})) - \mathbf{z}_0(\mathbf{x}) - \mathbf{\Phi}_z(\mathbf{x})\boldsymbol{\lambda} \right\|_2^2 \quad (4)$$

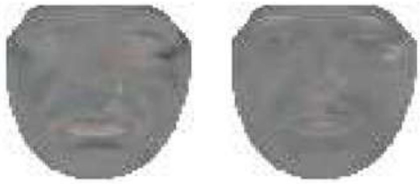


$\mathbf{s}_0$ $\mathbf{s}_1$ $\mathbf{s}_2$

$\mathbf{s}_3$ $\mathbf{s}_4$

a) Shape model

$\mathbf{z}_0(\mathbf{x})$ $\mathbf{z}_1(\mathbf{x})$ $\mathbf{z}_2(\mathbf{x})$



$\mathbf{z}_3(\mathbf{x})$ $\mathbf{z}_4(\mathbf{x})$

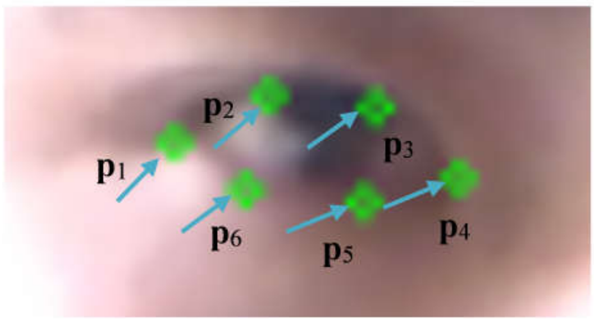
b) Appearance model

Figure 1. The linear shape model and appearance variation of an independent AAMs

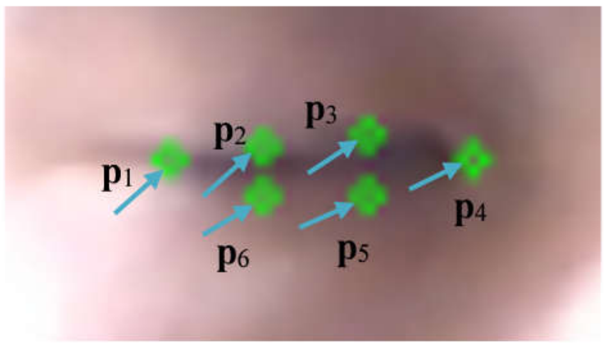
Solving the optimal function in Eq. (4) is considered difficult due to the existence of many local minima. Therefore, to solve this problem, the AOM variant model is proposed. The AOM is modeled by the correlation of the experimental image with the learned image model. The optimal estimates are given by:

$$\{\tilde{\mathbf{c}}, \tilde{\boldsymbol{\lambda}}\} = \underset{\{\mathbf{c},\boldsymbol{\lambda}\}}{\operatorname{argmax}} \frac{\mathbf{z}(\mathbf{x})[\mathbf{c}]^T \mathbf{\Phi}_z \boldsymbol{\lambda}}{\left\| \mathbf{\Phi}_z \boldsymbol{\lambda} \right\|_2} \quad (5)$$

3. EXPERIMENT SYSTEM



a) Landmarks in eye-opening state



b) Landmarks in the state of eyes closed

Figure 2. Landmarks around the subject's eye position

The proposed method was used to create annotations for facial landmark presented in Section 2. We proceed to create a training database used from Multi-PIE [10]. The Multi-PIE face database contains around 750,000 images of 337 subjects under 15 view points, 19 illumination conditions and displaying 6 different facial expressions. For every video frame in real time, the eye landmarks are detected. The eye ratio between height and width of the eye is given by:

$$W = \frac{\|\mathbf{p}_2 - \mathbf{p}_6\|_2 + \|\mathbf{p}_3 - \mathbf{p}_5\|_2}{2\|\mathbf{p}_1 - \mathbf{p}_4\|_2} \qquad (6)$$

where $\mathbf{p}_j$ (j = 1...6) are the 2D landmark locations. $\|\cdot\|_2$ denote the normed vector space.

We investigated eye ratio in 3 cases: (i) Change in light intensity: from well-lit to low-light conditions; (ii) Subjects wear glasses and do not wear glasses; (iii) Subject's posture changes. Blink time was also studied to determine the duration of 5 frames of the video. If an eye ratio W < 0.25 occurs, it is considered a blink. Based on real-time signals, during the investigation period, we tried to avoid this blinking phenomenon. Survey time is displayed in 10s (each 0.1s get a frame result): eye-opening time 5s, eye closing time 5s. Figs. 3-5 describes eye ratio in light intensity conditions ranging from well-lit, lack of light, low light.

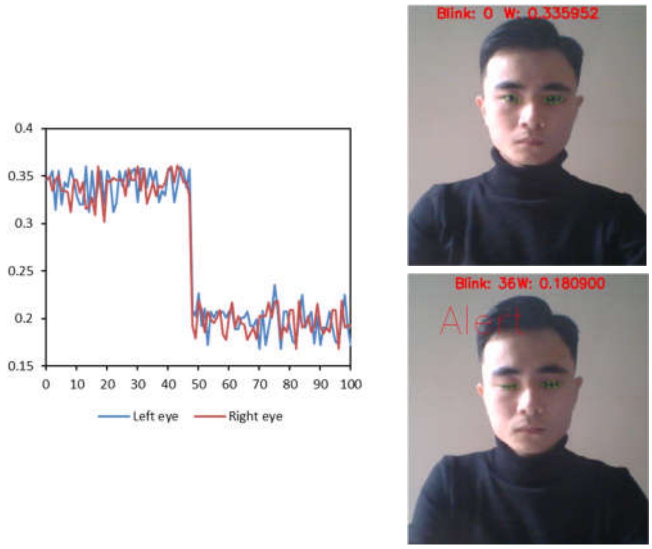


Figure 3. Ratio of eyes in well-lit conditions

From Figs. 3-5 we see that:

i) In well-lit conditions when eyes are open, the value W is from 0.32 to 0.34 (left eye) and from 0.29 to 0.34 (right eye). When eyes are closed, the value W drops suddenly to a value lower than 0.25 indicating eye closure, the value ranges from 0.17 to 0.24 (left eye) and from 0.17 to 0.23 (right eye). Due to the well-lit conditions, fluctuations in eye ratio are low in real-time.

ii) In lach of light conditions when eyes are open, the value W is from 0.28 to 0.37 (left eye) and from 0.27 to 0.35 (right eye). With eyes closed the value W is from 0.16 to 0.23 (left eye) and from 0.16 to 0.23 (right eye). Despite lack of light conditions, the ability to detect objects and track the ratio is still maintained. However, the amplitude of oscillation is larger than that of the well-lit case. It shows that the object detection ability of the algorithm is reduced in low light conditions.

iii) In low-light conditions, the algorithm's recognition ability is significantly reduced, it is almost impossible to track objects in real-time.

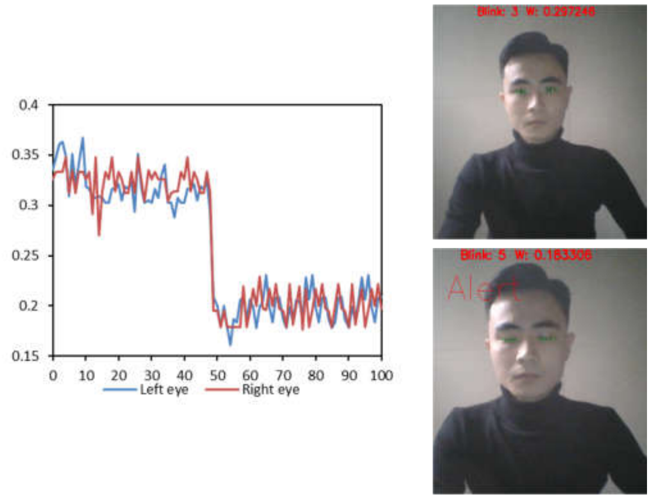


Figure 4. Ratio of eyes in lack of light conditions

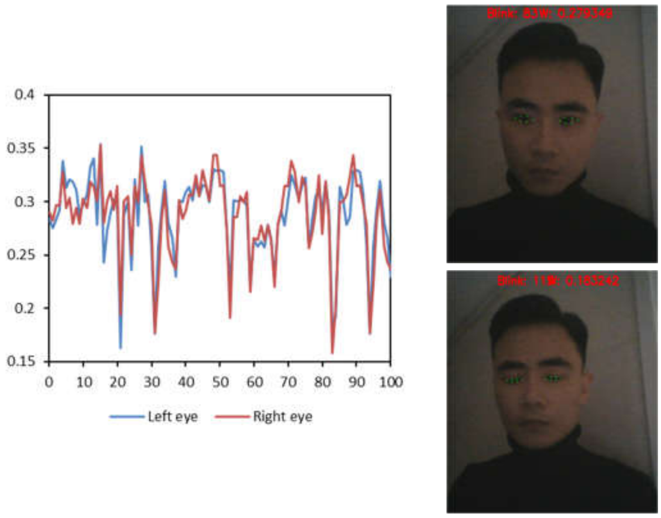


Figure 5. Ratio of eyes in low-light conditions

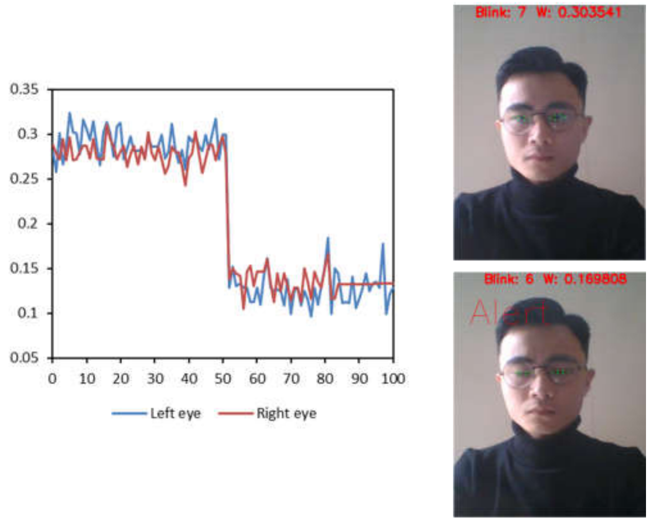


Figure 6. Ratio of eyes in glasses condition

When subject wears glasses, from Fig. 6 we see that the eye ratio has an amplitude ranging from 0.26 to 0.33 (eyes open) and from 0.1 to 0.18 (eyes closed). The ratio of eyes convergence is reduced but not large, indicating that the subject wearing glasses has little influence and the algorithm still responds well.

Program results are controlled in the Multi-PIE database. Therefore, we continue to conduct the investigation in different poses of the subject. Posture estimation is the process of determining an object's position at real coordinates based on the object's coordinates on the camera. The author's team proceeds to set the object's deflection angles with respect to the camera's projection angle. The survey results are depicted in Figures 7 and 8.

a) Eyes open (γ=10deg)

b) Eyes closed (γ=10deg)

c) Eyes open (γ=20deg)

d) Eyes closed (γ=20deg)

e) Eyes open (γ=30deg)

f) Eyes closed (γ=30deg)

g) Eyes open (γ=40deg)

h) Eyes closed (γ=40deg)

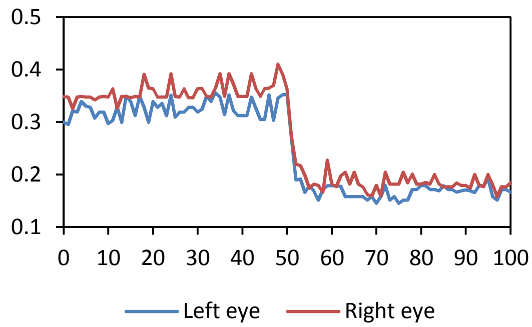
Figure 7. Subject's changing poses

From Figure 7 and Figure 8, we see that:

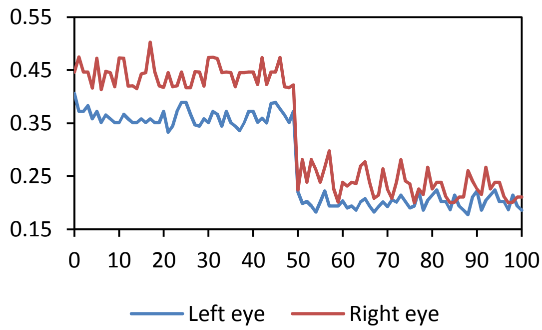
i) In a well-lit environment, the eye area has no confounding factors but the eye ratio is measured at an angle of deviation from the straight line of sight, so the oscillation graph gradually increases with the deviation angle.

ii) The difference between the ratio of the eyes in the left and right eyes occurs in the same frame. Without loss of generality, when the subject turns to the right, the left eye will be in the effective detection area, so the distance calculation algorithm will calculate accurately, and for the right eye, it will be farther away, so the ability of detection and calculation will decrease with the deviation angle.

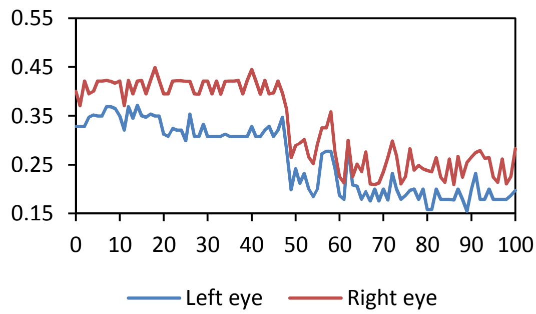
iii) When γ = 10deg ÷ 20deg, the evaluation effect is still quite good. However, when γ = 30deg the oscillation is increased significantly, the results of eye closure are not accurate. When γ = 40deg, both eye-opening and eye-closed states are unresponsive, the algorithm reaches a critical threshold that no longer guarantees the accuracy.



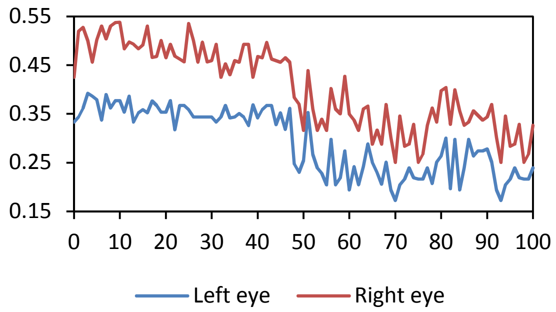
a) Ratio of eyes when γ=10deg



b) Ratio of eyes when γ=20deg



c) Ratio of eyes when γ=30deg



d) Ratio of eyes when γ=40deg

Figure 8. Ratio of eyes when changing the subject's position

4. CONCLUSION

From the discussions in sections 2 and 3 of the paper. This study had the following results:

i) This study has built a mathematical model from models and database data sets to determine facial feature landmarks.

ii) Building a function to determine eye ratio value to monitor eye-opening behavior in real-time. Since then, the study has investigated the sensitivity of the model under different conditions.

iii) When the subject's eyes are closed for a specified period of time the algorithm will issue a warning.

iv) The results of this research are expected to be able to classify drivers who are drowsy or not in real-time by observing the driver's eye activity. In addition, the results of the study can be used to detect the necessary light thresholds to ensure a stable system and can be used in system security issues. These issues will be presented in our next study in the near future.

REFERENCES

[1]. T. Li, W. Hou, F. Lyu, Y. Lei, C. Xiao, 2016. *Face detection based on depth information using HOG-LBP*. Sixth International Conference on Instrumentation Measurement, Computer, Communication and Control (IMCCC), pp. 779-784.

[2]. Xudong Sun, Pengcheng Wu, Steven C.H. Hoi, 2018. *Face detection using deep learning: An improved faster RCNN approach*. Neurocomputing 299, pp. 42-50.

[3]. K. Zhang, Z. Zhang, Z. Li, Y. Qiao, 2016. *Joint face detection and alignment using multitask cascaded convolutional networks*. IEEE Signal Processing Letters, vol. 23, no. 10, pp. 1499-1503.

[4]. R. Qi, R.S. Jia, Q.C. Mao, H.M. Sun, L.Q. Zuo, 2019. *Face detection method based on cascaded convolutional networks*. IEEE Access, vol. 7, pp. 110740-110748.

[5]. Tereza Soukupova, Jan Cech, 2016. *Real-Time Eye Blink Detection using Facial Landmarks*. 21st Computer Vision Winter Workshop, pp. 1-8.

[6]. Guarin D.L., Yunusova Y., Taati B., Dusseldorp J.R., Mohan S., Tavares, J., van Veen M.M., Fortier E., Hadlock T.A., Jowett N., 2020. *Toward an automatic system for computer-aided assessment in facial palsy*. Facial Plast. Surg. Aesthetic Med., 22, pp. 42-49.

[7]. K. Malik, B. Smolka, 2014. *Eye blink detection using Local Binary Patterns.* International Conference on Multimedia Computing and Systems (ICMCS), pp. 385-390.

[8]. Iain Matthews, Simon Baker, 2004. *Active appearance models revisited*. Int. J. Comput. Vision 60, 2, pp. 135-164.

[9]. Sagonas C., Tzimiropoulos G., Zafeiriou S., Pantic M., 2013. *A Semi-automatic Methodology for Facial Landmark Annotation*. 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 896-903.

[10]. R. Gross, I. Matthews, J. Cohn, T. Kanade, S. Baker, 2010. *Multi-pie*. Image and Vision Computing, 28(5), pp.807-813.

THÔNG TIN TÁC GIẢ

Trần Ngọc Tiến[1], Trần Văn Nghĩa[1], Phạm Thành Trung[2], Nhữ Quý Thơ[1], Nguyễn Đức Minh[1]

[1]Trường Đại học Công nghiệp Hà Nội

[2]Trường Cao đẳng Kinh tế - Kỹ thuật Vĩnh Phúc