Pattern Recognition Letters 148 (2021) 45–53

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

A new DCT-PCM method for license plate number detection in drone images

Hamam Mokayed[a,*], Palaiahnakote Shivakumara[b,*], Hon Hock Woon[c], Mohan Kankanhalli[d], Tong Lu[e], Umapada Pal[f]

[a] Faculty of Computer Science, Electrical and Space Engineering, Luleå University of Technology, Luleå, Sweden
[b] Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia
[c] Advanced Informatics Lab, MIMOS Berhad, Kuala Lumpur, Malaysia
[d] School of Computing, National University of Singapore, Singapore
[e] National Key Lab for Novel Software Technology, Nanjing University, Nanjing, China
[f] Computer Vision and Pattern Recognition Unit, Indian Statistical Institute, Kolkata, India

ARTICLE INFO

Article history:
Received 21 October 2020
Revised 8 April 2021
Accepted 8 May 2021
Available online 23 May 2021

Keywords:
Discrete cosine transform
Phase congruency
License plate detection
Scene text detection
Deep learning
Drone images

ABSTRACT

License plate number detection in drone images is a complex problem because the images are generally captured at oblique angles and pose several challenges like perspective distortion, non-uniform illumination effect, degradations, blur, occlusion, loss of visibility etc. Unlike, most existing methods that focus on images captured by orthogonal direction (head-on), the proposed work focuses on drone text images. Inspired by the Phase Congruency Model (PCM), which is invariant to non-uniform illuminations, contrast variations, geometric transformation and to some extent to distortion, we explore the combination of DCT and PCM (DCT-PCM) for detecting license plate number text in drone images. Motivated by the strong discriminative power of deep learning models, the proposed method exploits fully connected neural networks for eliminating false positives to achieve better detection results. Furthermore, the proposed work constructs working model that fits for real environment. To evaluate the proposed method, we use our own dataset captured by drones and benchmark license plate datasets, namely, Medialab for experimentation. We also demonstrate the effectiveness of the proposed method on benchmark natural scene text detection datasets, namely, SVT, MSRA-TD-500, ICDAR 2017 MLT and Total-Text.

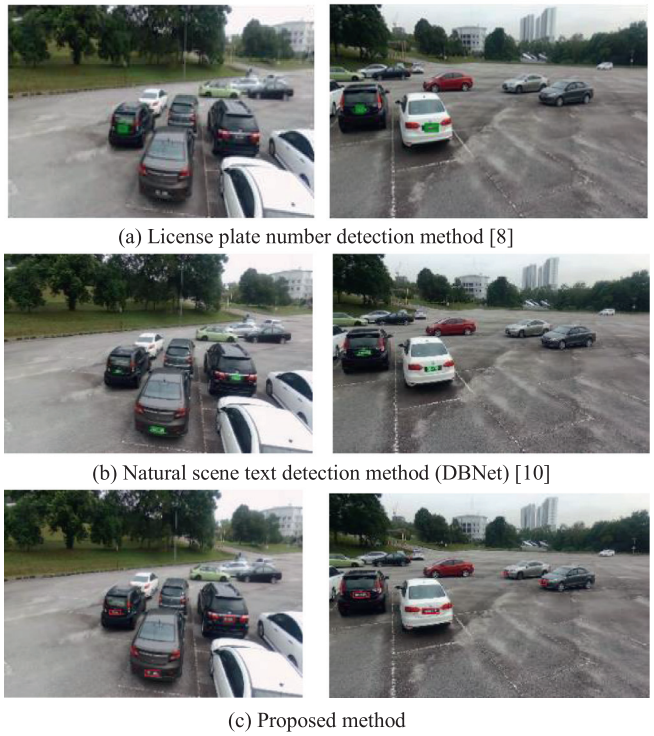
1. Introduction

For most countries, it is obvious that density of car movements increases exponentially due to urbanization, employment and living style changes. This leads to congestion in traffic and it causes severe problem of traffic monitoring and other security related issues [1–3]. For instance, monitoring exhibitions and parades or processions, where it is necessary to use drones for controlling the vehicle movements according to situations and locations. In these situations, drone captures vehicle images with outdoor scene as background. As a result, the complexity of the car number detection increases. In addition, the proposed car number detection method can also be used for other surveillance applications, such as toll fee collection, vehicle detection, which violates traffic signal and monitoring speed of vehicles.

For license plate number detection, there are several methods developed based on powerful deep learning models in the past for addressing the challenges of low contrast complex background, font size variations etc. License plate number detection methods from dirty plates and images are affected by different weather conditions [4]. License plate detection from the images affected by illumination of lights and auto setting of camera [5,6] and complex background [7] and high efficient vehicles and license plate detection [8]. However, these techniques do not work well for the images captured by drones. This is due to the effect of variation in height distance and oblique angel compared to the images captured in head on-direction. Since drone captures usually many vehicles in a single image as shown in Fig. 1, the focus of the camera spreads across vehicles. Therefore, as height distance changes, the number of vehicles in a single image also changes and this affects the quality of each license plate number.

* Corresponding authors.
E-mail addresses: hamam.mokayed@ltu.se (H. Mokayed), shiva@um.edu.my (P. Shivakumara), hockwoon.hon@mimos.my (H.H. Woon), mohan@comp.nus.edu.sg (M. Kankanhalli), lutong@nju.edu.cn (T. Lu), umapada@isical.ac.in (U. Pal).

https://doi.org/10.1016/j.patrec.2021.05.002
0167-8655/



(a) License plate number detection method [8]

(b) Natural scene text detection method (DBNet) [10]

(c) Proposed method

Fig. 1. Illustrating the challenges of text detection in drone images. Detected texts are shown in red and green boxes.

In other words, a single image can contain license plate numbers affected by different qualities. This makes license plate detection task in drone images much more complex and challenging compared to normal images captured by head-on direction, which includes text detection in natural scene images.

Similarly, if we consider license plate number as text in natural scene images, we can find sophisticated methods for text detection based on deep learning models [9,10]. However, since these methods are developed for the images captured by orthogonal direction, the approaches may not have ability to cope with the challenges of drone images mentioned-above. It is evident from the Fig. 1(a) and (b), where the state-of-the art methods of license plate number [8] and natural scene text detection [10] report poor results for the drone images, respectively. On the other hand, the proposed method detects license plate number of almost all the vehicles accurately as shown in Fig. 1(c). To the best of our knowledge, this is the first work towards addressing the challenges of license plate detection in drone images. It is noted that the proposed method is effective for text detection in natural scene images also. This is the key contribution of the proposed method and is different from the stat-of-the-art.

2. Related work

In the past, several methods are proposed for license plate number and text detection. We review these methods of license plate number detection as well as natural scene text detection.

To grasp the overall picture of existing work quickly, we summarize the existing methods for natural scene and license plate detection in Table 1. It is noted from Table 1 that the existing methods used deep learning and hand-crafted features for both license plate number and text detection in images. When we look at the objective of the approaches, none of the methods aims drone images. In addition, the limitations of the approaches show that the existing methods are good for addressing challenges posed by the images captured in orthogonal direction but not the challenges of drone text images.

In order to cope with the challenges posed by drone images, we propose a novel method by exploring the combination of DCT and Phase Congruency Model (DCT-PCM). Motivated by the method [25] where it is noted from DCT transform that high frequency coefficients captures information about the noisy pixels, low frequency coefficients capture information about background pixels, and middle coefficients contain information about the edge pixels. In case of license plate number detection in the complex environment, edges play a key role in representing text. The coefficients that contain edge pixels information exhibit coherence property such as uniform magnitude, orientation independent angle symmetry, contrast, uneven illumination, script etc. Therefore, though the image is affected by adverse factors caused by the drone, the DCT coefficients help us to separate edge information of the text in the images irrespective of aforementioned challenges. To handle the above challenges efficiently, inspired by the method [26] that used Phase Congruency Model for detecting objects in phytoplankton images, we propose to extract the coherence property of text edge pixels in DCT domain. The phase congruency model helps in extracting the pixels that share uniform direction and magnitude, which are key properties of edges representing text. The key contribution of this work is to propose the combination of DCT and Phase Congruency Model (DCT-PCM) for addressing challenges of drone images license plate number detection.

3. Proposed method

In the case of text image, the edges that represent text share similar properties. This is not true for the edges that represent non-text. This is because text usually written by the same color and it has homogeneous background compared to non-text where one expects non-uniform color and heterogeneous background. When there are variations in the color and background, the edges may not exhibit the similar properties. To extract such observation, the combination of DCT-PCM is introduced. The DCT coefficients help us for extracting the properties of edges that represent text through direction and magnitude. Then the direction and magnitude information are used for estimating phase congruency. Therefore, this combination results in edges that share the same direction and magnitude, which outputs text edges. In addition, the phase angle and magnitude are invariant to contrast, resolution variations and un-even illumination effect. This is the motivation behind to propose phase congruency using DCT in this work. In summary, when the fine details are degraded due to the effect of variation in height distance and angle, the combination helps us to enhance the fine details, which represent edges of text.

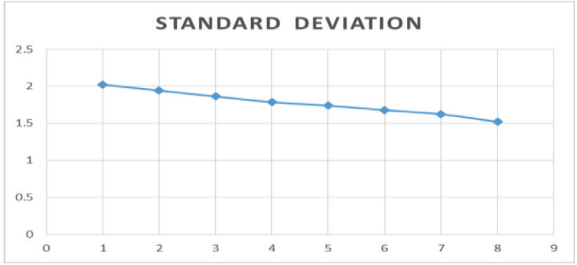
To show the phase congruency is invariant to different causes, which generally appear in drone text images, we estimate phase congruency for samples chosen randomly from the drone-text image dataset. The sample cropped images are shown in Fig. 2(a), where the average of standard deviation of phase congruency is almost same for the different causes as shown in Fig. 2(b). Therefore, it is expected that the phase congruency values for text pixels must be high compared to the background pixels. To classify the text pixels, we apply K-means clustering with K=2, resulting in text cluster, which are called candidate pixels. Since K-means clustering is unsupervised, the proposed method considers the cluster that gives high mean as text cluster. Due to complex nature of the problem, some of the background pixels may misclassify as candidate pixels.

To overcome this problem, we propose a new clustering approach for eight neighbors of each candidate pixel to discard false candidate pixels. It works based on the intuition that the neighbor pixels of the text share almost the same values and other pixels do

Table 1
The details of existing methods for natural scene and license plate detection are listed.

Methods	Idea	Objective	Limitation	Application	Drone
Shivakumara et al. [7]	Handcrafted features	Word detection	Generalization	Keyword spotting	No
Peng et al. [8]	Deep learning	LPD	Vehicle positon	License plate	No
Bartz et al. [9]	Deep learning	End-to-end scene text recognition	Learning parameters	Scene	No
Liao et al. [10]	Deep learning	Scene text detection	Single characters	Scene	No
Shi et al. [11]	Deep learning	Multi-oriented text detection	Multi-lingual	Scene	No
Tian et al. [12]	Tracking based	Text detection	Still images	Video	No
Liao et al. [13]	Deep learning	Small font, irregular size text	Learning parameters	Scene	No
Xu et al. [14]	Deep learning	Irregular sized text	Post processing	Scene	No
Rong et al. [15]	Deep learning	Action in the images	Segmentation	Scene	No
Musil et al. [16]	Features	Object detection	Specific application	Scene	No
Baek et al. [17]	Deep learning	Scene text detection	Irregular sized character	Scene	No
Li et al. [18]	Deep learning	Scene text detection	Generalizability	Scene	No
Panahi et al. [4]	Features	Robust LPD	Particular plates	License plate	No
Laroca et al. [19]	Deep learning	Recognition	Segmentation	License plate	No
Raghunandan et al. [20]	Features	Enhancement	Preprocessing	License plate	No
Moreno et al. [1]	Features	Distance variation	Orthogonal direction	License pate	No
Liu et al. [2]	Features + CNN	LPD	Preprocessing	License plate	No
Li et al. [21]	Deep learning	End-to-end recognition	Particular dataset	License plate	No
Shemarry et al. [22]	Features	LPD	Robustness	License plate	No
Chen et al. [23]	Features	Vehicle and LPD	Expensive	License plate	No
Peker [24]	Deep learning	LPD	Language specific	License plate	No

(a) Sample license plate numbers affected by different adverse effects. Sample-1: 480×360 resolution, Sample-2: Rotated by 30-degree angle (clockwise), Sample-3: low contrast, Sample-4: 240×140 resolution, Sample-5: Non-uniform illumination effect, Sample-6: Distortion due to blur with orientation, Sample-7: Non-uniform illumination effect with certain orientation, Sample-8: 360×240 resolution with certain orientation.



(b) Uniform standard deviation of Phase congruency for license plate images affected by different distortion

Fig. 2. Illustrations of the nvariance of the Phase Congruency values to geometric transformation and its adverse effects to some extent.

not. As a result, one can expect linearity relationship between the neighbors of text pixel while non-linearity relationship between neighbors of non-text pixel. The pixels, which satisfy linear relationship, are called seed pixels.

3.1. DCT phase congruency model for candidate pixel detection

The proposed model employs the DCT for each sliding window of size 8×8 of the pixels in the image to get complex transform as defined in Eq. (1),

$$F_{cc}(u,v) = C_1C_2\sum_{m=0}^{P-1}\sum_{n=0}^{q-1} f(m,n)\cos[\frac{\pi u}{P}(m+0.5)]\cos[\frac{\pi v}{q}(n+0.5)] \quad (1)$$

$$F_{cs}(u,v) = C_1C_2\sum_{m=0}^{P-1}\sum_{n=0}^{q-1} f(m,n)\cos[\frac{\pi u}{P}(m+0.5)]\sin[\frac{\pi v}{q}(n+0.5)] \quad (2)$$

$$F_{sc}(u,v) = C_1C_2\sum_{m=0}^{P-1}\sum_{n=0}^{q-1} f(m,n)\sin[\frac{\pi u}{P}(m+0.5)]\cos[\frac{\pi v}{q}(n+0.5)] \quad (3)$$

$$F_{ss}(u,v) = C_1C_2\sum_{m=0}^{P-1}\sum_{n=0}^{q-1} f(m,n)\sin[\frac{\pi u}{P}(m+0.5)]\sin[\frac{\pi v}{q}(n+0.5)] \quad (4)$$

where: 0< u < 7 and 0<v <7, C1 and C2 are the constants

$$C_i = \{\begin{matrix}\sqrt{\frac{2}{P}}, & 1\le u\le 7 \\ \sqrt{\frac{1}{P}}, & u=0\end{matrix} \quad i=1,2 \quad (5)$$

By combining the four real transforms as:

$$[F_{cc}(u,v) - F_{ss}(u,v)] - j[F_{cs}(u,v) + F_{sc}(u,v)] \quad (6)$$

The transform will be as follows:

$$F(u,v) = C_1C_2\sum_{m=0}^{P-1}\sum_{n=0}^{q-1} f(m,n)e^{-j\frac{\pi u}{P}(m+0.5)}e^{-j\frac{\pi v}{q}(n+0.5)} \quad (7)$$

Both the phase (ϕ) and Amplitude (A) are calculated for each pixel of the image from Eq. (7) as follows

$$\phi_n = atan2(\frac{F_i(u,v)}{F_r(u,v)}) \quad (8)$$

$$A_n = (\sqrt{F_r(u,v)^2 + F_i(u,v)^2}) \quad (9)$$

(a) Input image (b) DCT-PCM (c) Candidate pixels

Fig. 3. The phase congruency for candidate pixels detection

The phase angle ϕ and Amplitude A of each location in the signal are used for deriving phase congruency as defined in Eq. (10). A phase congruency function in terms of the discreet cosine expansion of the signal at some location x defined as:

$$PC(x) = max_{\bar{\emptyset}(x)\in[0,2\pi]} \frac{\sum_n A_n \cos(\emptyset_n(x) - \bar{\emptyset}(x))]}{\sum_n A_n} \tag{10}$$

where A_n represents the amplitude, and $\emptyset_n(x)$ represents the local phase of Cosine component at position x. Value $\bar{\emptyset}(x)$ that maximizes this equation is the amplitude weighted mean local phase angle of all the transform terms at the point being considered. Since the energy is proportional to the cosine of the deviation of phase angle, the sine of phase difference in addition to the cosine is proposed to construct a more robust phase deviation measure as defined in Eq. (11).

$$PC(x) = \frac{\sum_n [A_n(x)(\cos(\emptyset_n(x) - \bar{\emptyset}(x)) - |\sin(\emptyset(x) - \bar{\emptyset}(x))|) - T]}{\sum_n A_n(x) + \in} \tag{11}$$

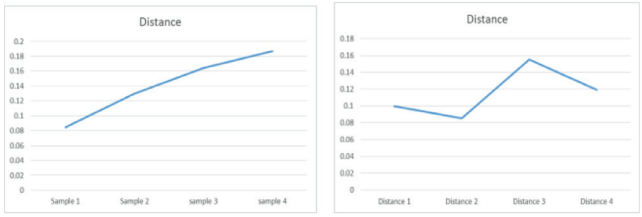
here, T is the threshold for noise compensation which is the mean noise plus multiple, k, σ_R

$$T = \mu_R - k\sigma_R$$

where k is in the range between 2 to 3, $\mu_R$is the mean of the Rayleigh distribution and $\sigma_R$is the variance of it, ε is the constant to avoid division by zero.

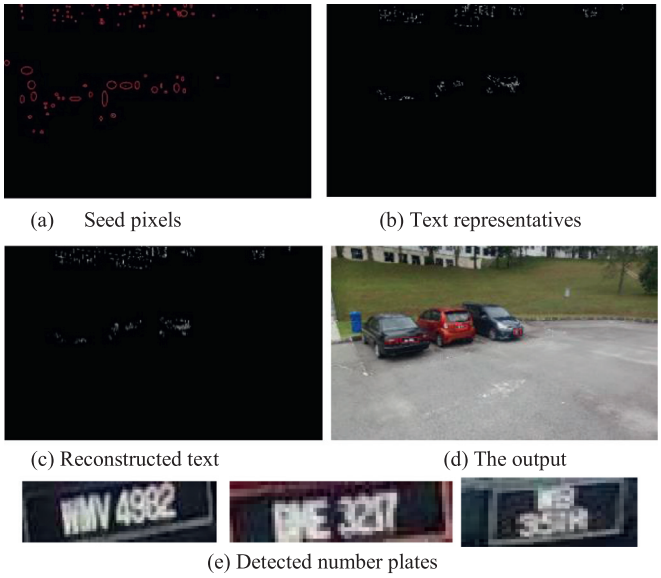
The effect of DCT-PCM for the input image in Fig. 3(a) can be seen in Fig. 3(b), where one can notice that the pixels that represents edge are highlighted and other pixels are suppressed. This indicates that the pixels, which represent text have high values and other pixels have low values. Therefore, the proposed model uses K-means with K=2 clustering for classifying the pixels which represents text as text cluster. This results in candidate pixels as shown in Fig. 3(c). It is observed from Fig. 3(c) that it contains not only the pixels, which represents text but also non-text.

To overcome the problem of false text candidate pixels, the proposed work introduces a new clustering approach, which considers 8 neighbors of each candidate pixels. For the set of 8 neighborhood elements, the method finds the smallest element out of 8 elements. Then the other 7 elements are compared with the chosen minimum element to find the next smallest element out of 7 elements, which results in the first pair cluster. For the rest of next 6 elements, the method repeats the same step to get the second pair cluster. After the first and second clusters, the set contains only four elements. When the same process repeats for the set consists of four elements, it gives the third and fourth pair clusters. For each pair cluster, we calculate absolute difference, which gives four such difference values as defined in Eqs. (6) and (7). For the set of elements representing text, the absolute difference of four pair clusters has almost the same. Otherwise, the difference of each pair clusters increases linearly. For the set of elements representing non-text, one cannot expect linear relationship between the difference values. It is illustrated in Fig. 4, where it can be seen that the line of candidate text pixels exhibits linearity behavior and



Candidate text pixels False text pixels

Fig. 4. Illustrating linearity and non-linearity check for candidate and false text pixels



(a) Seed pixels (b) Text representatives

(c) Reconstructed text (d) The output

(e) Detected number plates

Fig. 5. Examples of linearity and non-linearity for obtaining text representatives for each seed pixel.

the line of false text pixels exhibits non-linearity behavior. For example, let S be the set contains 0.23, 0.48. 0.67. 0.19. 0.43. 0.66. 0.13 and 0.31. The absolute difference values of the first, second, third and fourth pair clusters are 0.60, 0.07, 0.34, 0.00, respectively. When we look at the difference values, there is a big difference between the first and second differences compared to second-third, while there is a small difference between the second-third compared to third-fourth. This defines non-linearity behavior. Therefore, the pixel is considered as a false text candidate. If the set of element representing text, it consists of 0.24, 0.22, 0.18 0.39. 0.37. 0.31, 0.88 and 0.59. The difference values are as follows, 0.03, 0.06, 0.17 and 0.28. When we look at difference values, it is small but the difference increases gradually as cluster pair number increases, and hence it defines linearity behavior. Thus, it is a text candidate pixel. This results in seed pixels marked by red color as shown in Fig. 5(a).

$$slope = (sum_{xy} - sum_x * y_{mean})/(sum_{xx} - sum_x * x_{mean}) \tag{6}$$

where

$$sum_x = \sum_{i=0}^{i=3} xDist_i$$

$$sum_y = \sum_{i=0}^{i=3} yDist_i$$

$$sum_{xy} = \sum_{i=0}^{i=3} xDist_i * yDist_i$$

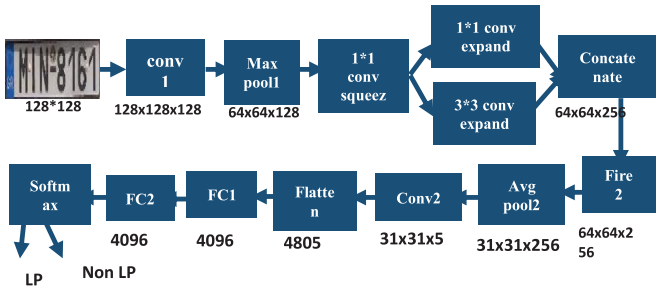


Fig. 6. Example of Deep learning architecture that works for false positive elimination. Here NP denotes License plate and Non LP denotes false positives.

$$sum_{xx} = \sum_{i=0}^{i=3} xDist_i * xDist_i$$

$$\text{If}(\text{diff} < \beta) -> \text{nominated seed pixels} \quad (7)$$

where

$$diff = \sum_{i=0}^{i=3} abs(yDist_i - (slope * xDist_i + (y_{mean} - slope * x_{mean}))$$

The β in Eq. (7) is small threshold determined experimentally, which allows some tolerance for checking linearity and non-linearity.

For each seed pixel, the proposed system retrieves the edge components from Canny edge image of the input image, which corresponds to seed pixels as shown in Fig. 5(b), where we can see restored edges for corresponding to seed pixels, which we called text representatives. The text representatives are used to restore the missing text information by referring to Canny edge images of the input image as shown in Fig. 5(c). For this, the proposed system uses nearest neighbor criteria based on distance and direction of the text information. Further, the proposed system fixes bounding boxes for the reconstructed results as shown in Fig. 5(d). For fixing bounding boxes, our method considers boundary points of every character component and uses those boundary points to determine regression line based on direction and character shape information. This process fixes bounding boxes for any orientation (arbitrarily-oriented text). Fig. 5(e) shows license plate number detection with bounding boxes.

3.2. Deep learning model for text detection in drone images

Motivated by the strong discriminative power of deep learning model [27], we use the same for removing false positives such that license plate number detection performance improves. In this work, the presence of multiple cars in a single image makes false positive elimination more difficult because the car parts and color may share the same features of text. Therefore, we propose to use a fully connected neural network as shown in Fig. 6 for removing false positives.

This architecture begins with a standalone convolution layer (conv1), followed by a max-pooling with a stride of 2. Next the output of max-pooling (max pool1) is processed by two layers of fire layer (fire1 and fire2). A fire layer includes a squeeze convolution layer (which has only 1×1 filters) and then feeding into an expand layer that has a mix of 1×1 and 3×3 convolution filters. Finally, the result from conv2 is flattened and is passed through 3 fully-connected layers and softmax function. The output result from this architecture is 2 classes (which are noise or license plate). The input image size is 128×128×3 and training process is using Adam optimization with training rate 0.001. The total number of images used in the database training is 250k plates/250k noises). The overall accuracy is 98.6427% and the average inferencing speed per image frame using a CPU of i7-8700K is 32 ms.

Fig. 7. License plate detection of the proposed system.

Fig. 8. Drone used for capturing images at different height distances and angles.

The proposed approach is a combination of the feature extraction and deep learning model and it integrates the merits of both to achieve the good results by handling the challenges of license plate number images. The output normalization for all the classifiers is done with soft equation as defined in Eqs. (8)–(10).

$$S(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}} \quad (8)$$

Loss computation for each training data is done using cross entropy loss as defined in Eq. (9)

$$H(P,\ Q) = \mathbb{E}_{x\sim P}[-\log Q(x)] \quad (9)$$

At the end of each iteration, the average loss is calculated as defined in Eq. (10)

$$average\ loss = \frac{\sum loss}{n} \quad (10)$$

The effect of deep learning can be seen in Fig. 7, where it is noted that the false positives are eliminated, resulting in license plate number detection in drone images.

3.3. Constructing a working model

Since the aim of the proposed work is to develop practical approach for license plate number detection in drone images in real environment, we use the drones shown in Fig. 8 for experimentation. The model of the drone is Parrot BeBop 2, which consists of 14 Megapixel "Fisheye" Camera with 3-Axis Stabilization. The built-in camera can record up to 1080p video as well as capture photos at 4096×3072 resolution. Notably, the camera uses a 3-axis electronic image stabilization system to avoid the added weight and power consumption of a mechanical gimbal. The FreeFlight 3 app enables live monitoring of the camera image along with flight telemetry overlay. Using the app's touch interface, we can even digitally pan the image across the camera's 180° field of view. We use car park areas of our research institute for experiments. The video is converted to images of size 1280×760 through Parrot FreeFlight Pro application.

The framework of working model is shown in Fig. 9. The video captured by drone is processed in LPR Server, which consists of the

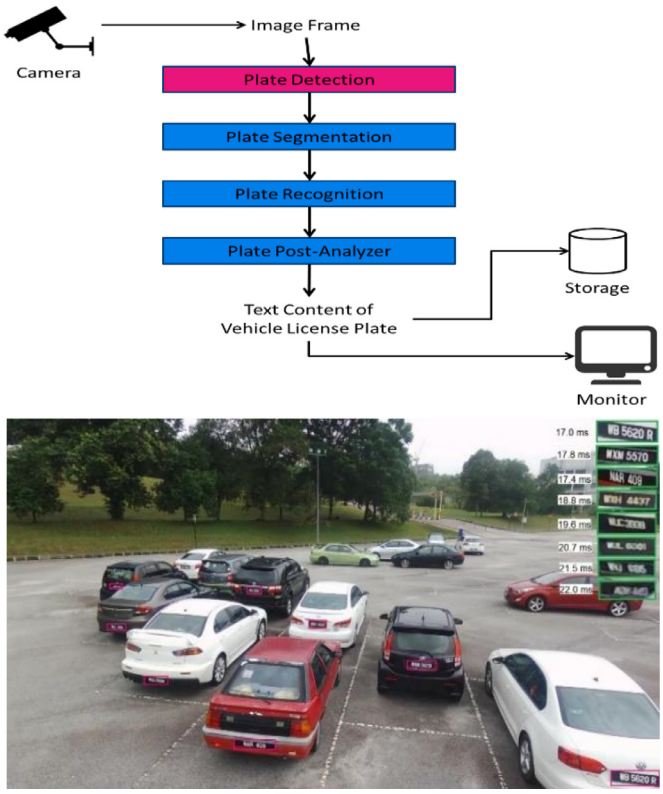


Fig. 9. Frame work of the proposed live prototype system.

following components, Video Analysis (VA) Video Analysis Command Service, LPR Engine and Live Notifications Engine. The VA acquires video streams to convert to the sequence of images for processing such as text detection. It also takes care of network connectivity to video sources and it monitors the process to ensure that the camera source is still accessible. The LPR Engine manages the issues of memory management during frame processing. It is like bridge between commands (image processing) and service (configures/starts/stops the LPR Engine). For example, the engine accepts the frame and loads into a shared memory for license plate processing. It also sends a message of successful detection and completion of the task.

The video analysis command service sends signal for action of all the process, such as capture images, perform license plate detection, etc. and hence it behaves as a controller. The component of live notification engine notifies external clients by sending the coordinates, snapshot of the detected plate, and the full frame. The "live networking" helps us to save the whole process in the form of database according to events requested by the clients. For sample image output of the proposed prototype system received by clients (users of the research) is shown in Fig. 9 where we can see the license plate numbers are detected for multiple cars in the image. As soon as the system detects license plate, it displays the same over the image.

4. Experimental results

For evaluating the proposed system, we have created our own dataset because there is no standard dataset available in the literature for license plate number detection in drone images. Our dataset includes the images captured at different timing, such as 8:00 am, 12:00 and 5:00 pm to get data of different car density from the open car parking area in our campus. In addition, the dataset includes the images captured from varying height distance from 1-3, 3-5, and more than 7 meters at different angles. This gives total 1000 images for experimentation. To test the effectiveness of the proposed method and to ensure that it works well for the images captured by orthogonal direction, we consider the Medialab [28] benchmark dataset, which contains 680 images affected by small font, different distance variations from orthogonal directions and poor quality due to defocusing. In total, $1000 + 680 = 1680$ images are considered for evaluating the proposed system for text detection in license plate images.

As discussed in the introduction section, the images captured by drones affected by two additional causes, namely, change in character structure and unpredictable quality of images due to variations on height distance and oblique angle compared to the images captured by head on-direction, which includes the natural scene images of the standard datasets. Therefore, one can say that natural scene images captured in head-on direction are a sub-set of drone images. As noted in [28], license plate number detection is a part of scene text detection. This is because license plate number is a part of text. When the proposed method works well for complex drone images, it should work well for natural scene images captured in head-on direction. Therefore, we conduct experiments on natural scene images of benchmark datasets and compared the proposed method with the state-of-the art natural scene text detection methods.

We consider benchmark datasets of natural scene text detection, namely, SVT, MSRATD-500, ICDAR 2017 MLT and Total-text. A short description of these datasets is given here. **SVT (Street View Text)** [29]: Since the dataset focus on capturing building street and shop names in urban area, one can expect the effect of little oblique angle as drone images. It provides 350 samples for testing. Therefore, this is a relevant dataset for evaluating the methods.**MSRA-TD500** [17] provides 300 training images and 200 test images for evaluation. The dataset includes multi-oriented texts of English and Chinese languages. The ground truth is available at line level. **ICDAR 2017 MLTdataset** [17] provides multilingual text images containing 7200 training images, 1800 validation images and 9000 testing images. The dataset consists of images of 9 languages in arbitrary orientations. This dataset is used for testing the ability of multi-lingual text detection. **Total-Text** [30] has 1255 images for training and 300 for testing. The dataset includes curved text with low resolution, low contrast and complex backgrounds. This is more complex dataset than other natural scene text datasets. Note that the images of these dataset are captured by orthogonal direction but not drone.

To compute the performance of the proposed and existing methods, we follow standard instructions and measures mentioned in [30]. The standard measures, namely, Recall (r), Precision (p) and F-measure (f) are used for all the experiments in this work. If the best match which is calculated using ground truth and bounding boxes given by the methods is larger than 0.5 (50%) then the detected license plate/text is considered as correct count (true positive). Otherwise, it is considered as false negative.

To show the usefulness of the proposed method, we implemented several state-of-the-art methods for comparative study. Laroca et al. [21] proposed license plate detection based on YOLO detector. Peker [24] explored tensorflow object detection networks for license plate number detection. Peng et al. [8] used end-to-end secondary network for license plate detection. The above license plate number detection methods addresses the several challenges similar to drone images and they used different deep learning models for achieving better results. Although, the existing methods use deep learning models for better performance, the methods may not report satisfactory results for drone images. In the same way, to show that the natural scene text detection method does not cope with the challenges of drone images, we have used the latest and state-of-the-art methods for comparative study. Bartz

Table 2
Analyzing the effectiveness of the key steps of the proposed method using our dataset.

Experiments	The key steps	R	P	F
(i)	Proposed method without DCT with PCM	76.2	81.1	78.5
(ii)	Proposed method without PCM with DCT	78.6	83.4	80.9
(iii)	Proposed method without Canny edge detection	77.1	79.3	78.1
(iv)	Proposed method without deep learning	71.1	84.7	77.3
(v)	Proposed Method (Baseline)	83.2	86.2	83.9

Our dataset Medialab dataset

Fig. 10. Examples of qualitative results of the proposed system.

Table 3
Performance of the proposed and existing methods on our and Medialab datasets.

Methods	Our dataset			Medialab dataset		
Measures	R	P	F	R	P	F
SEE [9]	50.0	60.0	54.5	72.0	70.0	71.0
OpenAlpr [19]	72.0	68.0	69.9	76.0	75.0	75.5
Peker [24]	71.7	62.7	66.9	75.4	71.9	73.6
Peng et al. [8]	86.7	56.9	68.7	83.4	66.8	74.2
Proposed	83.2	86.2	83.9	83.1	80.1	81.1

et al. [9] proposed towards semi-supervised end-to-end scene text recognition, which is called SEE. Baek et al. [17] for Character Region Awareness for Text detection (CRAFT), Li et al. [18] for Shape Robust Text Detection with Progressive Scale Expansion Network (PSENet) and Liao et al. [10] for Real time scene text detection with differentiable Binarization (DBNet). However, it is noted that all the above-mentioned existing methods designed for the images captured by head-on direction.

4.1. Ablation study

It is noted that the proposed method consists of several key steps for license plate number detection in drone images, namely, use of DCT coefficients for enhancing fine details, Phase congruency for sharpening fine details, Canny edge detector to restore character components using seed points, deep learning for false positive elimination, etc. To show that the steps are effective and contribute for achieving the best results, we conduct the following experiments using our dataset to calculate the measures. Experiment (i) The phase angle and magnitudes are estimated using FFT instead of DCT for obtaining PCM for the input image. (ii) PCM is used for text candidate detection using K-means clustering. (iii) In this case, the proposed work considers DCT coefficient matrix directly without PCM for detecting text candidates. (iv)The proposed method uses seed points without restoring character components with the help of Canny edge detector for text detection. (v) In this experiment, the deep learning used for removing false positives is replaced by rules with geometrical features. The results of all the five experiments are reported in Table 2, where it can be seen that the results of all the experiments report low results compared to those of the proposed method (baseline). This shows that the steps which are missed for text detection are effective for achieving the best results for license plate number detection in drone images. At the same time, we can also conclude that all the steps contribute equally for the best results.

4.2. Experiments for license plate detection in drone and orthogonal direction images

Qualitative results of the proposed method on drone images and the images of Medialab datasets are shown in Fig. 10, where it can be seen that our method detects license plate number accurately in both the images. Therefore, we can conclude that the proposed system has ability to handle challenges of both drone images and normal images captured by orthogonal direction. Quantitative results of the proposed and existing methods for our and Medialab datasets are reported in Table 3, where it is evident that the proposed system is the best in terms of precision and F-measure compared to existing methods including natural scene text detection method [9]. This shows that the existing methods [8,9,19,24] are inadequate to beat the proposed method. However, the method [8] is the best at Recall for both drone and Medialab datasets compared to the proposed method because it is developed for handling distorted license plate images, such as low light and scaled images. But it produces more number of false positives for both drone and Medialab images and hence the precision is lower than the proposed method. The reason for the poor results of the exiting methods is that the methods suffer from inherent limitations, such as language specific features, constraints on scope, shape and vehicles positions etc. On other hand, since the proposed method involves the combination of DCT and Phase congruency, which are invariant to challenges posed by drone text images, the proposed method is the best.

4.3. Experiments for text detection in natural scene images

Sample qualitative results of the proposed system for SVT, MSRA-TD 5000, ICDAR 2017 MLT and Total-Text datasets are shown in Fig. 11(a)–(d), respectively. It is observed from Fig. 11(a)–(d) that the proposed system is good for street view text, which may be affected by little oblique angle like drones, horizontal text, multi-oriented text, multi-lingual text and curved text detection in natural scene images. Quantitative results of the proposed and existing methods on SVT, MSRA-TD500, ICDAR 2017 MLT and Total-text dataset are reported in Table 4, where it can be seen that the proposed system achieves consistent results especially Recall irrespective of complexities of the datasets. On the other hand, the existing methods are not consistent as the approaches are the best for one dataset and worst for another dataset. Thus, we can confirm that the existing methods are not robust to different datasets that pose different challenges.

Our method scores the best precision, F-measure for SVT dataset, Recall for MSRA-TD500 and ICDAR 2017 MLT datasets. But for Total-text dataset, the proposed method is not best compared to existing methods but it scores competitive results. For Total-text dataset, the CRAFT is the best at Recall and DBNet is the best at Precision and F-measure. Since the target of the proposed

Table 4
Performance of the proposed and existing methods on benchmark natural scene datasets.

Methods	SVT			MSRATD-500			ICDAR 2017 MLT			Total Text		
Measures	R	P	F	R	P	F	R	P	F	R	P	F
CRAFT [17]	87.2	73.1	79.5	78.2	88.2	82.9	80.6	68.2	73.9	87.6	79.9	83.6
PSENet [18]	54.0	69.8	60.8	52.0	85.9	64.5	75.3	69.2	72.2	84.0	75.2	79.6
DBNet [10]	62.2	72.5	67.0	79.2	91.5	84.9	67.9	83.1	74.7	82.5	87.1	84.7
Proposed	86.5	75.4	80.5	88.4	76.3	81.2	82.2	71.2	73.5	85.1	80.4	82.3

Table 5
Recall of the proposed system for different heights distance with angles on our drone image dataset.

Varying distances with Angle	Recall –(different distances)	Recall-(different angles)	APT in ms
1-3 meter and 0- ±10 angles	95.2	91.3	18.45
3-5 meter and ±10 - ±20 angles	85.4	87.1	19.63
5-7 meter and ±20 - ±40 angles	79.8	80.3	20.22

(a) SVT (b) MSRA TD-500

(c) ICDAR 2017 MLT (d) Total-Text

Fig. 11. Text detection of the proposed system for the natural scene images of different benchmark datasets.

method is to tackle the challenges of both drone and normal license plate/natural scene images, it achieves consistent and competitive results for different datasets while the existing methods developed specifically for naturals scene images, the method score the best results for Total-Text compared to the proposed method. Hence, we can conclude that the proposed method is robust to different datasets, complexities, type of text such as license plate and natural scene text, and drone, normal images.

4.4. Evaluating the robustness of the proposed working model

To show effectiveness of the proposed system, we choose 500 images randomly from our datasets with different distance and angle variations as mentioned in Table 4. For experiments, we choose the images by varying only distance and fixing angle as one set and varying angles while fixing distance as a second set. We calculate recall and Average Processing Time (APT) for each experiment. The results reported in Table 5shows that the Recall is promising for the images of different height distances and angles. The APT for the different experiments reported in Table 4 shows that the proposed system is fast and it is fit for real time applications. It is also noted from Table 4 that as distance increases, the APT also increases because the images cover large area with many cars. This leads to poor results and requires more computation. Overall, the proposed system scores similar or better results for both text in drone and natural scene text datasets.

5. Conclusion and future work

In this work, we have proposed a novel method for license plate number detection in drone as well as normal images. The proposed method uses the combination of DCT and Phase Congruency Model in novel way for candidate pixel detection in the images. The combination extracts the edges that represent text irrespective of poses created oblique angles, height distance variations and defocus. A new clustering approach is also used for eliminating false candidate pixels. Furthermore, a fully connected neural network has been explored for eliminating false positives to improve detection performance. Experimental results on our dataset of drone images as well as benchmark datasets of orthogonal direction show that the proposed approach outperforms the existing methods. Our next target is to explore temporal information for improving performance of the proposed system at different situations in real environment. To the best of our knowledge, our work on license plate number detection in drone images is the first of its kind. The code, dataset and ground truth will be released to publicly to support reproducibility.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors acknowledge the support from the grant, FRGS, Ministry of Higher Education, Malaysia, FP104-2020 for this work. And also, the support from the Natural Science Foundation of China associated with Grants, 61672273, 61832008 and the Science Foundation for Distinguished Young Scholars of Jiangsu associated with the Grant BK20160021.

References

[1] M.M. Moreno, I.G. Diaz, F.D.D. Maria, Efficient scale-adaptive license plate detection system, IEEE Trans. Intell. Transp. Syst. 20 (2019) 2109–2121.
[2] C. Liu, F. Chang, Hybrid cascade structure for license plate detection in large visual surveillance scenes, IEEE Trans. Intell. Transp. Syst. 20 (2019) 2122–2135.
[3] L. Xie, W. Ahmad, L. Jin, Y. Liu, S. Zhang, A new CNN based method for multi-directional car license plate detection, IEEE Trans. Intell. Transp. Syst. 19 (2018) 507–517.
[4] R. Panahi, I. Gholampour, Accurate detection and recognition of dirty vehicle plate numbers for high speed applications, IEEE Trans. Intell. Transp. Syst. 18 (2017) 767–779.

[5] Y. Su, J.Y. Lin, C.C.J. Kuo, A model based approach to camera's auto exposure control, J. Vis. Commun. Represent. 35 (2016) 122–129.
[6] Y. Su, C.C.J. Kuo, Fast and robust camera's auto exposure control using convex of cancave model, in: Proc. ICCE, 2015, pp. 13–14.
[7] P. Shivakumara, S. Roy, H.A. Jalab, R.W. Ibrahim, U. Pal, T. Lu, V. Khare, A.W.B.A. Wahab, Fractional means based method for multi-oriented keyword spotting in video/scene/license plate images, Expert Syst. Appl. 118 (2019) 1–19.
[8] Y. Peng, H. Li, Z. Qian, A new end to end secondary network for high efficient vehicles and license plate detection, in: Proc. ICSGEA, 2019, pp. 6–9.
[9] C. Bartz, H. Yang, C. Meinel, SEE: towards semi-supervised end-to-end scene text recognition, in: Proc. AAAI, 2018, pp. 6674–6681.
[10] M. Liao, Z. Wan, C. Yao, K. Chen, X. Bai, Real time scene text deteciton with differential binarization, in: Proc. AAAI, 2020.
[11] B. Shi, M. Yang, X. Wang, P. Lyu, C. Yao, X. Bai, ASTER: an attentional scene text recognizer with flexible rectification, IEEE Trans. Pattern Anal. Mach. Intell. 41 (9) (2018) 2035–2048.
[12] S. Tian, X.C. Yin, Y. Su, H.W. Hao, A unified framework for tracking based text detection and recognition from web videos, IEEE Trans. Pattern Anal. Mach. Intell. 40 (2018) 542–554.
[13] M. Liao, B. Shi, X. Bai, Textbox++:a single shot oriented scene text detector, IEEE Trans. Image Process. 8 (2018) 3676–3690.
[14] Y. Xu, Y. Wang, W. Zhou, Y. Wang, Z. Yang, X. Bai, TextField: learning a deep direction field for irregular scene text detection, IEEE Trans. Image Process. 28 (2019) 5566–5579.
[15] X. Rong, C. Yi, Y. Tian, Unamiguous scene text segmentation with referring expression comprehension, IEEE Trans. Image Process. 29 (2020) 591–601.
[16] P. Musil, R. Juranek, M. Musli, P. Zemcik, Cascaded stripe memory engine for multi-scale object detection in FPGA, IEEE Trans. Circ. Syst. Video Technol. 30 (2020) 1–13.
[17] Y. Baek, B. Lee, D. Han, S. Yun, H. Lee, Character Region Awareness for Text Detection, in: Proc. CVPR, 2019, pp. 9357–9366.
[18] X. Li, W. Wang, W. Hou, R.Z. Liu, T. Lu, J. Yang, Shape robust text detection with progressive scale expansion network, in: Proc. CVPR, 2019, pp. 928–9337.
[19] R. Laroca, E. Severo, L.A. Zanloresnsi, L.S. Oliveira, G.R. Goncalves, W.R. Schwartx, D. Menotti, A robust real time automatic license plate recognition based on the YOLO detector, in: Proc. IJCNN, 2018.
[20] K.S. Raghunandan, P. Shivakumara, H.A. Jalab, R.W. Ibrahim, G.H. Kumar, U. Pal, T. Lu, Riesz fractional based model for enhancing license plate detection and recognition, IEEE Trans. Circ. Syst. Video Trans. 28 (2018) 2276–2288.
[21] H. Li, P. Wang, C. Shen, Toward end to end car license plate detection and recognition with deep neural networks, IEEE Trans. Intell. Transp. Syst. 20 (2019) 1126–1136.
[22] M.S.A. Shemarry, Y. Li, S. Abdulla, An efficient texture descriptor for the detectio of license plates from vehicle images in difficult conditions, IEEE Trans. Intell. Transp. Syst. 21 (2019) 1–12.
[23] S.L. Chen, C. Yang, J.W. Ma, F. Chen, X.C. Yin, Simultaneous end to end vehicle and license plate detection with multi branch attention neural networks, IEEE Trans. Intell. Transp. Syst. 21 (2019) 1–10.
[24] M. Peker, Comparision of tesnsof flow object detection networks for license plate localization, in: Proc. GPECOM, 2019, pp. 101–105.
[25] D. Raví, M. Bober, G. Farinella, M. Guarnera, S. Battiato, Semantic segmentation of images exploiting DCT based features and random forest, Pattern Recognit. 52 (2016) 260–273.
[26] A. Verikas, A. Gelzinis, M. Bacauskiene, I. Olenin, E. Vaicukynas, Phase congruency based detection of circular objects applied to analysis of phytoplankton images, Patten Recognit. 45 (2012) 1659–1670.
[27] Y. Su, C.C.J. Kuo, On extended long shor-term and dependent bidirectional recurrent neural network, Neurocomputing 356 (2019) 151–161.
[28] A. Zamberletti, I. Gallo, L. Noce, Augmented text character proposals and convolutional neural networks for text spotting from scene images, in: Proc. ACPR, 2015, pp. 196–200.
[29] T.Q. Phan, P. Shivakumara, C.L. Tan, Detecting text in the real world, in: Proc. ACM MM, 2012, pp. 765–768.
[30] C.K. Ch'ng, C.S. Chan, Total-text: a comprehensive dataset for scene text detection and recognition, in: Proc. ICDAR, 2017, pp. 935–942.