RESEARCH   **Open Access**

Vision-based patient identification recognition based on image content analysis and support vector machine for medical information system

Guo-Shiang Lin[1*], Sin-Kuo Chai[2], Hsiang-Min Li[2] and Jen-Yung Lin[3]

* Correspondence: gslin@ncnt.edu. tw
[1]Department of Computer Science and Information Engineering, National Chin-Yi University of Technology, Taichung, Taiwan
Full list of author information is available at the end of the article

Abstract

In this paper, a vision-based patient identification recognition system based on image content analysis and support vector machine is proposed for medical information system, especially in dermatology. This proposed system is composed of three parts: pre-processing, candidate region detection, and digit recognition. To consider the efficiency of the proposed scheme, image normalization is performed. The color information is used to identify camera-captured screen images. In the pre-processing part, the effect of noise in captured screen images is reduced by a bilateral filter. The color and spatial information is used to initially and roughly locate the candidate region. To reduce the skew effect, a skew correction algorithm based on the Hough transform is developed. A template matching algorithm is used to find special symbols for locating the region of interest (ROI). For digit segmentation, digits are segmented in the ROI based on the vertical projection and adaptive thresholding. For the digit recognition, some features are measured from each digit segment and a classifier based on the support vector machine is applied to recognize digits.

The experiment's results show that the proposed system could effectively not only use color information to distinguish the captured screen images from the skin images but also detect the ROIs. After the digit segmentation, the accuracy rates of digit recognition are 98.4% and 94.2% for the proposed system and the Tesseract Optical Character Recognition (OCR) software, respectively. These results demonstrate that the proposed system outperforms the Tesseract OCR software in terms of the accuracy rate of digit recognition.

Keywords: Patient identification recognition, Digit recognition, Support vector machine

Lin et al. EURASIP Journal on Advances in Signal Processing (2020) 2020:27

Page 2 of 15

1 Introduction

Over the past few years, digital cameras have become almost ubiquitous and their popularity for capturing the scenery is unchallenged. In fact, mobile devices with camera have become an important image acquisition device in real applications of image analysis and computer vision. For example, a mobile device with license recognition technology can make the police easily identify stolen cars. In hospitals, physicians can use a device with the bar code reader to easily access patients' medical records. In fact, physicians can also use a vision-based device with digit recognition technology to easily recognize the patient identification information (PII) and then find the corresponding medical records. For instance, physicians often capture patients' skin images to record their current situations in dermatology. To help physicians recognize the PII, the screen images with PII are also captured. Figure 1 illustrates the camera-captured images in dermatology. The top and bottom columns of Fig. 1 are the screen and skin images, respectively. During the diagnosis period, physicians can recognize the PII from the screen image, the PII can be used to find these skin images from medical information system (MIS), and then physicians can understand the condition of patient's skin from the corresponding skin images. Moreover, the PII can effectively archive patients' skin images into MIS for reducing physicians' working load in dermatology. This means that effectively recognizing PII from camera-captured images is a significant procedure of computer aided diagnosis in dermatology. To this end, it motivates us to develop an efficient vision-based scheme to recognize PII from camera-captured images.

So far, there are no vision-based systems for recognizing PII captured from camera-captured images. A patient identification recognition system often has two main parts: region of interest (ROI) detection and digit recognition. There are many articles about ROI detection. In fact, the definition of ROI may change in different applications [1, 2]. For instance, ROI may be the region containing car license plates, personal information, or the suspects in video surveillance
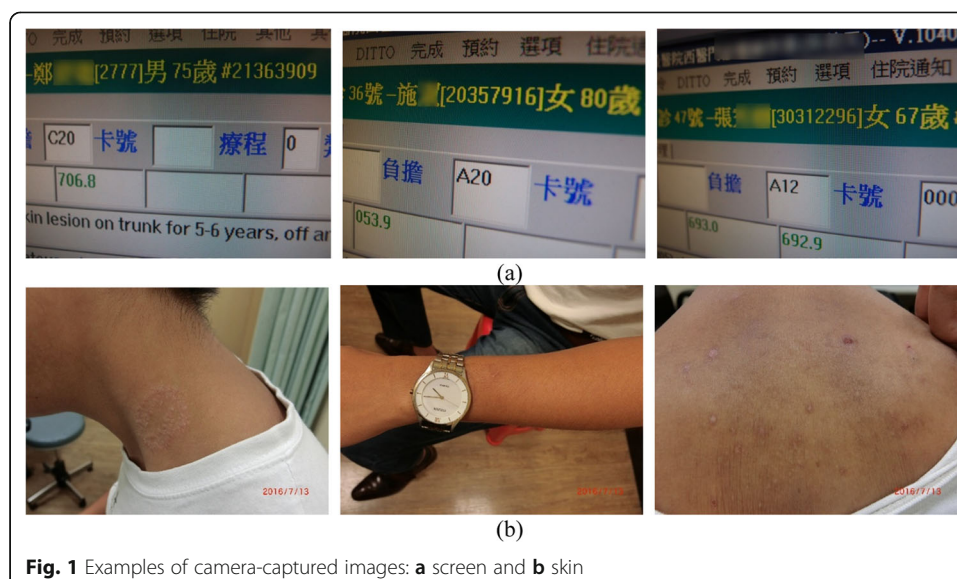

Fig. 1 Examples of camera-captured images: **a** screen and **b** skin

applications. In MIS, a ROI can be a region including the PII or tumors. In [1], the ROI containing tumors is located by removing the background part to raise the efficiency of tumor detection. As shown in captured screen images of Fig. 1, the PII is on the green region. This means that the region containing the PII can be considered as ROI within captured screen images in dermatology. The other one is digit recognition. Digit recognition can be exploited to recognize postcode, license plate, street number, etc. Until now, there are some existing methods for handwritten digit recognition [3–5]. An intuitive technique is template matching [6]. Unfortunately, the template matching approach often fails when the skew effect exists in unconstrained images. According to [3, 4], a feature-based handwritten digit recognition method often has two parts: feature extraction and digit classification. In [3], some feature extraction approaches were described for handwritten character recognition. For example, some features such as gradient histogram [3] or projection histograms [5] can be measured from an image for digit recognition. For digit classification, some classification techniques such as support vector machine (SVM) and artificial neural network can be used as a classifier. However, as shown in Fig. 1, the digits are the printed style. Since printed digits often have regular shape, it is easier to recognize printed digits compared with handwritten digits. Therefore, a digit recognition algorithm based on machine learning technology can be designed to recognize printed digits in the proposed system.

Because it is the main target to obtain the PII in camera-captured images for MIS by detecting and analyzing ROIs, some effects should be considered for devising the proposed system. Due to image acquisition, captured images are often not the same size and the skew effect often occurs. It is difficult to automatically recognize PII from images captured under the uncontrolled environments. For example, the visual quality of camera-captured images may be poor due to the following factors: motion blurring, skew effect, image resolution, noise, etc. As shown in Fig. 1 a, the skew effect and noise exist. It is no doubt that these factors do increase the difficulty of automatically recognizing the PII from the camera-captured images. To automatically recognize the PII from the camera-captured images, the skew effect and noise should be eliminated as much as possible. This means that the pre-processing including the skew correction is necessary in the proposed scheme.

To effectively detect and analyze ROIs to obtain the PII from the camera-captured images, a system based on the content analysis and machine learning is developed. Figure 2 illustrates the block diagram of the proposed scheme. As shown in Fig. 2, the proposed scheme is composed of three parts: pre-processing, candidate region detection, and digit recognition. The pre-processing part is utilized to normalize the input image and then determine whether an input image is a screen image. The candidate region detection part is a coarse-to-

Fig. 2 Block diagram of the proposed system
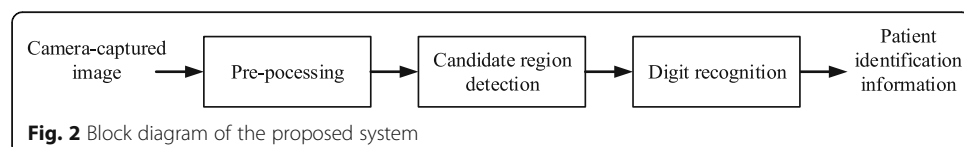
fine approach to locate the ROI for further analysis. The digit recognition part is used to segment and recognize each digit in the ROI. Each part is elaborated in the following.

The remainder of this paper is organized as follows. In section II, the pre-processing part is introduced; in section III, the candidate region detection is described; and in section IV, the digit recognition is elaborated. Section V details the results of experiments conducted to evaluate the system, and section VII contains some concluding remarks.

2 Pre-processing

In real applications, physicians can capture screen images and skin images by using a digital camera or smart phone in dermatology. To consider the efficiency of the proposed system, the pre-processing part is comprised of denoising, image normalization, and image classification. As well-known, a bilateral filter cannot only reduce the noise effect but also preserve the edge information in natural and depth images [7, 8]. To reduce the noise effect, a bilateral filter [7, 8] is the first step of pre-processing in the proposed system. The image normalization and image classification are described in the following.

A. Image normalization

In the actual environments, the spatial resolutions of captured images are often different due to different image acquisition devices. This means that different mobile devices with different resolutions may be used in real applications. It is often expected that the computational complexity of dealing with captured images with different spatial resolutions is high. To consider the efficiency of the proposed system, the input image is decimated to obtain a normalized version. The image normalization process is a down-sampling operation described as follows:

$$X^D = T_1(X), \tag{1}$$

where $T_1(\cdot)$ denotes the down-sampling operation and X^D is the normalized version. The down-sampling operation is a low-pass filter and then the high-frequency components are often attenuated in the decimated captured images [6].

B. Image classification

As shown in Fig. 1, there are more yellow, green, blue, and gray pixels in the screen images. However, because the skin images may contain not only the skin but also the patient's clothes and the background of the scene shot, the variance of color information is large in the captured skin images compared with the screen images. Therefore, the properties of color information can be used to identify the screen images for image classification.

The hue, saturation, value (HSV) color space corresponding to human perception has been used in the image analysis and computer vision [6, 9]. Here, the properties of color information in the HSV color space are analyzed for image

classification. The steps of distinguishing the screen images from the skin images are described as follows:

1. Normalize the input image:

To raise the efficiency of the proposed system, the size of the input image is normalized as 640×480 pixels.

2. Perform the color space transformation and count the numbers of yellow, green, and blue pixels in the HSV color space:

Since a screen image contains some yellow, green, and blue pixels, the color information can be exploited for determining whether an input is a screen image. The numbers of yellow, green, and blue pixels based on the HSV color space are determined in the following:

$$
M_1 Y(x,y) = \begin{cases} 1 & \text{if } \begin{array}{ll} 20 \le & I^H(x,y) \quad \le 50 \\ 140 \le & I^S(x,y) \quad \le 255 \end{array}, \\ 0 & \text{otherwise} \end{cases} \tag{2}
$$

$$
M_1^G(x,y) = \begin{cases} 1 & \text{if } \begin{array}{ll} 60 \le & I^H(x,y) \quad \le 90 \\ 150 \le & I^S(x,y) \quad \le 255 \end{array}, \\ 0 & \text{otherwise} \end{cases} \tag{3}
$$

$$
M_1^B(x,y) = \begin{cases} 1 & \text{if } \begin{array}{ll} 100 \le & I^H(x,y) \quad \le 130 \\ 200 \le & I^S(x,y) \quad \le 255 \end{array}, \\ 0 & \text{others} \end{cases} \tag{4}
$$

where I^H, I^S, and I^V represent the hue, saturation, and value components, respectively; N^Y, N^G, and N^B indicate the numbers of yellow, green, and blue pixels, respectively. (x, y) denotes the image coordinate. After the yellow, green, and blue pixels are determined based on Eqs. (2)–(4), N^Y, N^G, and N^B can be computed from M_1^Y, M_1^G, and M_1^B, respectively.

3. Determine whether the input is a screen image if the following conditions are satisfied: $N^Y \ge T^Y$, $N^G \ge T^G$, and $N^B \ge T^B$, where the thresholds are determined as $T^Y = 0.01$, $T^G = 0.33$, and $T^B = 0.0003$ based on our experiences.

After the image classification, the screen image can be identified and then analyzed for detecting the candidate region.

3 Candidate region detection

To effectively detect the candidate region in a screen image, a coarse-to-fine approach is developed. In the coarse-to-fine candidate region detection procedure, it contains three subparts: the initial ROI cropping, the skew correction, and the candidate region refinement. The initial ROI cropping locates the initial ROI region based on the spatial information. In fact, the screen may not be parallel to the image plane of the image acquisition device, so the camera-captured images often have the geometrical distortion and noise. Therefore, the skew correction is necessary for effectively detecting the candidate region. After the skew correction, a candidate region refinement algorithm is used to refine the candidate region for a further analysis.

 A. Initial ROI cropping

 Since the ROI is often located at the top part of a camera-captured screen image, the position of a candidate region is initialized on the top 1/3 region of the input image. In addition, as observed in Fig. 1a, the ROI often contains the yellow color information, so the yellow information can be extracted as a feature to roughly localize an area as the initial candidate region. Furthermore, in order to find the coordinates from the initial candidate region in the vertical axis, a horizontal projection operation is adopted. Thus, the procedure of undertaking the initial ROI cropping is described as follows:

 (A1) Initialize the position of a candidate region on the top 1/3 region of the input image.
 (A2) Select the yellow regions by using Eq. (2) and then perform horizontal projection [6] to obtain H_1 where H_1 is the output of the horizontal projection.
 (A3) Find the maximum value of H_1 and then estimate an adaptive threshold as follows:

$$Thd_1 = \alpha_1 \times \max\{H_1\}, \tag{5}$$

where $\max\{\cdot\}$ is the maximum operator and Thd_1 denotes the adaptive threshold. The parameter α_1 is set as 0.25.

 (A4) Binarize H_1 by using Thd_1 and then search for the top and button coordinates of the threshold H_1 to obtain the initial candidate region I.

 Then the initial candidate region I containing PII can be obtained.

 B. Skew correction

 There are many existing methods to conduct the skew angle detection for the scanned document images [10–13]. These existing skew angle detection methods can be categorized into several classes: the projection profile analysis, Hough transform, the nearest neighbor clustering, the cross-correlation, etc. [14–16]. Unfortunately, since the

characteristics of the screen images are different from those of document images, these existing methods [10–12] may not be suitable to estimate the skew angles of the captured screen images.

To reduce the impact of the skew effect on locating PII, a skew correction algorithm is used here. According to the structure property of PII in the screen images, the information of horizontal lines is useful for estimating the rotation angle. To let horizontal lines be detected easily, the impact of vertical lines on skew correction is decreased. The steps of the skew correction are listed below:

1. Binarize I by using the adaptive thresholding [17] to obtain the binary map B.
2. Examine the white points based on the following rules:

(a) $B_2Y(x, y)$ is white.
(b) $B_2Y(x, y{-}1)$, $B_2Y(x{-}1, y{-}1)$, and $B_2Y(x + 1, y + 1)$ are white.
(c) One of $B_2Y(x, y + 1)$, $B_2Y(x{-}1, y + 1)$, and $B_2Y(x{-}1, y)$ is black.

If the rules (a), (b), and (c) are satisfied, $B_2Y(x, y)$ is kept white; otherwise, $B_2Y(x, y)$ is set to be black.

3. Perform the Hough transform on B and then estimate the rotation angle θ of I.
4. Modify I based on the estimated angle θ.

After estimating the skew angle, the modified ROI I can be obtained by re-rotating I with the angle $(-\theta)$. It is expected that the skew effect can be reduced on I, and then I can be used for the further analysis.

C. Candidate region refinement

Due to the different situations for the image acquisition, the sizes of digits may not be the same among the captured screen images. However, if the background around digits could be removed as much as possible, it is easily to normalize the size of digits. Thus, the boundaries of ROI should be refined in the vertical and horizontal directions to eliminate the redundant background for better performance of the digit recognition.

The candidate region refinement modifies the vertical boundaries of ROI. The procedures similar to (A2) to (A5) are used to refine the top and button boundaries of the ROI. After refining the top and button boundaries of the ROI, the following procedure is to find the left and right boundaries of ROI.

To find the left and right boundaries of ROI, the candidate region refinement is developed based on the template matching. As shown in Fig. 1, the PII is bounded to the two specific symbols, "[" and "]." Therefore, the two templates containing the specific symbols can be used to find the left and right coordinates of ROI. The steps of refining the left and right boundaries of ROI are described as follows:

T1. Compute the ratio β of ROI and the right template TR as follows:

$$\beta = \frac{R^H}{TR^H},$$ (6)

where R^H and TR^H denote the heights of ROI and the right template, respectively. Here the right template is "]."

T2. Normalize the right template based on β.

T3. Compute the cross-correlation of ROI and the right template as follows:

$$S(u,v) = \frac{\sum_{x',y'}\left(TR(x',y') \times I_2^Y(u+x',v+y')\right)}{\sqrt{\sum_{x',y'}TR(x',y')^2 \times \sum_{x',y'}I_2^Y(u+x',v+y')^2}},$$ (7)

where TR is the right template and $S(u,v)$ represents the cross-correlation result at (u,v).

T4. Find the maximum value of cross-correlation and its position. The position is set as the coordinate of the right boundary of ROI.

T5. Repeat steps T1 to T4 to find the maximum value of cross-correlation and its position by using the left template. Here, the left template TL is "[." The position is set as the coordinate of the left boundary of ROI.

T6. Refine the ROI based on the coordinate of the right and left boundaries.

4 Digit recognition

Due to the skew effect, the template matching is not suitable in the proposed system. To recognize the PII from the ROI, there are two subparts: digit segmentation and digit recognition. For a ROI, the digit segmentation is used to segment each possible digit for the digit recognition. As mentioned in section I, recognizing each digit can be considered as a classification problem. This means that it is necessary to extract some useful features and construct a suitable classifier for the digit recognition. Moreover, as shown in Fig. 1, since the digits are the printed style, the proposed digit recognition algorithm should be devised for dealing with the printed digits in the camera-captured images. The two subparts will be introduced in detail in the following.

A. Digit segmentation

Before recognizing the digits in the ROI, digits need to be segmented from the ROI. As shown in Fig. 1a, the color of PII is different from the background. Based on the observation, the steps of the digit segmentation for a ROI are described as follows:

1. Perform histogram equalization [5].
2. Conduct Otsu thresholding [6].

3. Perform morphological opening operation [6].

4. Perform the vertical projection and measure an adaptive threshold Thd_2 as follows:

$$Thd_2 = \alpha_2 \times \ \max\{P^V(B_3)\}, \tag{8}$$

where B_3 denotes the binarization result after morphological opening operation and P^V is the vertical projection operation. The parameter α_2 is set as 0.10.

5. Detect the beginning and the end points of each digit based on the adaptive threshold Thd_2. The beginning and the end points of each digit can be expressed as below:

$$St_i = \ \min_u \left\{ V_1(u) > Thd \quad (9) \right.$$

$$En_i = \ \max_u \{V_1(u) > Thd_2\}, \tag{10}$$

where St_i and En_i represent the beginning and the end points of the i-th digit, respectively, $V_1 = P^V(B_3)$, and $\min\{\cdot\}$ is the minimum operator.

B. Digit classification

Since each digit segment may not have the same size, each digit segment should be normalized for the digit recognition. To recognize each detected digit segment, some useful features should be measured, and a suitable classifier should be designed. Hence, to classify each digit segment, there are two subparts: feature extraction and digit classifier.

As known, the structure among the printed digits is clear and different from each other. Figure 3 illustrates the feature extraction for recognizing each digit segment. For each digit segment, the steps of feature extraction are described as follows:

Fig. 3 Feature extraction for digit recognition: **a** four column regions and **b** eight row regions
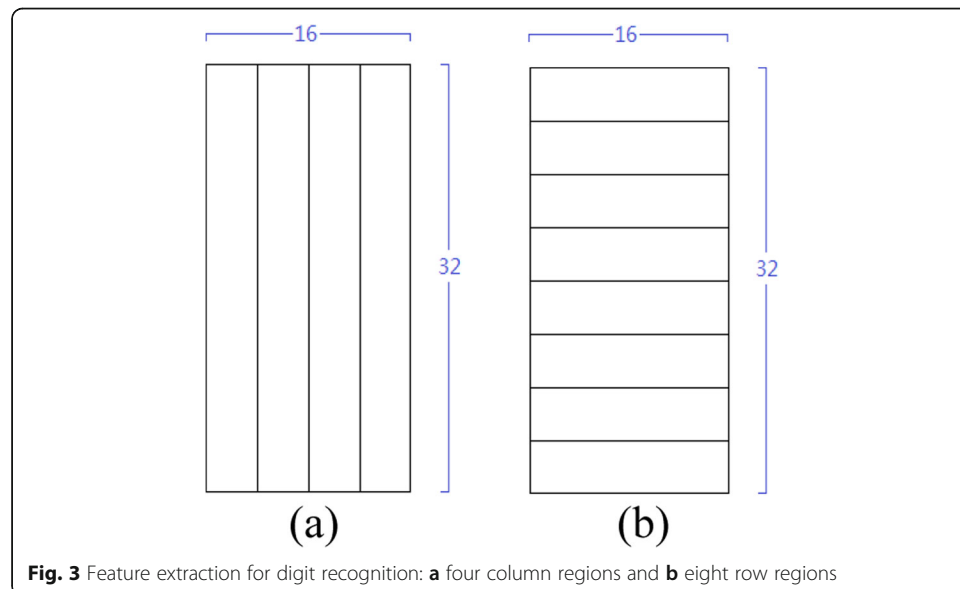
1. Normalize the digit segment into 16×32 pixels.
2. Divide the normalized digit segment into four column regions shown in Fig. 3a and compute features as follows:

$$F_i^1 = \sum_{x=4(i-1)}^{4i-1} \sum_{y=0}^{31} \Omega(x,y), \quad i = 1,2,3,4, \tag{11}$$

where Ω denotes digit segment and F_i^1 represents the features extracted from four column regions.

3. Divide the normalized digit segment into eight row regions shown in Fig. 3b and calculate features as follows:

$$F_i^2 = \sum_{x=0}^{15} \sum_{y=4(i-1)}^{4i-1} \Omega(x,y), \quad i = 1, ..., 8, \tag{12}$$

where F_i^2 represents the features extracted from the eight row regions.

It is expected that the structure of each digit measured by F_i^1 and F_i^2 is different from each other. Then F_i^1 and F_i^2 can be used as features for the digit recognition.

In fact, due to the lighting condition, the complex background, and the image resolution, the candidate regions may be noisy and then the noisy features often make the digit recognition difficult. In the machine learning field, a supervised learning algorithm, SVM, is popularly exploited in many applications such as character recognition, text categorization, and face recognition [18]. The basic idea of SVM is to simultaneously minimize the empirical classification error and maximize the geometric margin between different classes. To deal with the difficult classification problems, SVM maps the features to higher dimensional space and then a hyperplane which potentially provides a better classification performance in the mapped feature space could be found. Then one of SVM software, LibSVM [19], is used as a digit classifier here.

5 Experimental results

To evaluate the performance of the proposed system, 253 test images including 100 camera-captured screen images and 153 camera-captured skin images from dermatology are collected for testing. The frame sizes of test images are 1256×1920, 3648×2736, and 4000×3000 pixels. The radial basis function is chosen as the kernel function in the SVM classifier for digit classification. In this section, the performance index used in the evaluations is firstly described. Secondly, the proposed system is analyzed. Finally, the proposed system's performance is compared with that of the existing tool.

A. Performance index

To evaluate the proposed system in terms of its capability and preciseness, accuracy is considered as the performance index [14–16, 20]. The accuracy rate is a traditional criterion that is widely used to measure the performance in many applications such as shot change detection and object detection. The accuracy rate is determined as follows:

$$\text{Accuracy} = \frac{N_{TP} + N_{TN}}{N_{TP} + N_{TN} + N_F + N_M}, \tag{13}$$

where N_C, N_M, and N_F are the numbers of correct detections, missed detections, and false alarms, respectively; $(N_C + N_M)$ is the total number of true objects; and $(N_C + N_F)$ is the total number of detected objects. $N_{TP} = N_C + N_M$ and N_{TN} denote the number of truth positive and true negative, respectively. Theoretically, it is assumed that the higher the accuracy can be, the better the detection rate of the proposed system can be achieved.

B. Analysis of the proposed scheme

The proposed scheme is analyzed in the following. Figure 4 a illustrates three samples of camera-captured screen images. Note that due to the privacy issue, the patients' names are partially blurred. As observed in Fig. 4a, the PII digits among these captured screen images are with different size and located in different relative position of the captured images. The skew effect also exists. The different size, relative position, and skew effect result from that the position and the view angle of the camera relative to the screen are not fixed during image acquisition. The unconstrained image acquisition conditions do raise the difficulty of locating and recognizing PII in the camera-captured screen images. That is why the proposed scheme is developed.

B.1 Pre-processing

There are 253 test images including 100 screen images and 153 skin images for testing. All of the screen images can be correctly identified. The experimental results show that the proposed system based on the color information can effectively distinguish the screen images from the skin images.

B.2 Initial ROI cropping

Figure 4 b illustrates the results of the initial ROI cropping. Compared with Fig. 4a, the initial candidate regions shown in Fig. 4b can be effectively detected from the input
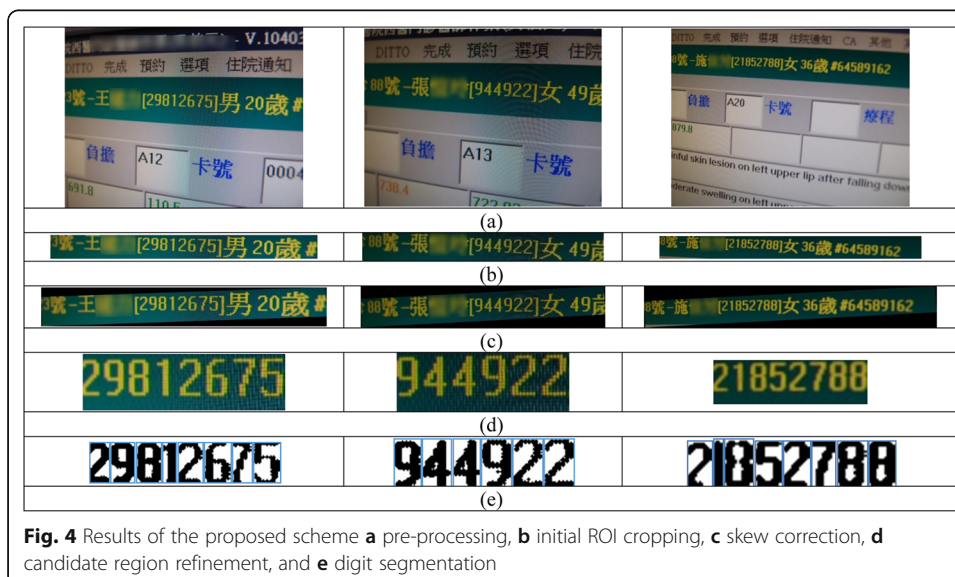

Fig. 4 Results of the proposed scheme **a** pre-processing, **b** initial ROI cropping, **c** skew correction, **d** candidate region refinement, and **e** digit segmentation

images. These experimental results show that the proposed system can effectively remove the background, the initial ROIs contain the PII and the search space can be effectively reduced for locating PII. In addition, although the initial ROIs can be roughly detected, the skew effect still exists in the initial ROIs shown in Fig. 4b. That is why the skew correction is needed in the proposed system.

B.3 Skew correction

Figure 4 c illustrates the results of the skew correction. As shown in Fig. 4c, the skew effect is visually reduced compared with Fig. 4b. The experimental results show that the proposed skew correction algorithm can estimate the rotation angles and then decrease the skew effect effectively.

To evaluate the performance of the skew correction in the proposed system, 20 images with the skew effect are simulated for testing. Compared with the ground truth, the average, minimum, and maximum errors of the proposed skew correction algorithm are 0.35°, 0.05°, and 1.09°, respectively. The experimental results demonstrate that the proposed skew correction algorithm can effectively reduce the skew effect.

B.4 Candidate region refinement

Figure 4 d shows the result of the proposed candidate region refinement algorithm. As we can see in Fig. 4d, the ROIs can be correctly relocated by using the proposed candidate region refinement algorithm and only PII exists in the refined ROIs. In addition, to evaluate the performance of the proposed candidate region refinement algorithm, the error rate is defined as the position difference of ROI between the ground truth and the detected version. The ground truth is obtained by manually locating the ROI in the captured screen images. For the test images, the maximum and minimum values of ROIs are 203 and 95 pixels in width. The maximum and minimum values of ROIs are 57 and 29 pixels in height. The average values of error rates are 1.03 and 1.02 pixels in the horizontal and vertical directions after refining the candidate regions by using the proposed algorithm. The experimental results show that the proposed candidate region refinement algorithm can efficiently refine the boundaries of the ROIs.

C. Digit recognition

Here, the performance of digit recognition in the proposed system is analyzed.

C.1 Digit segmentation

Figure 4 e shows the results of digit segmentation. As shown in Fig. 4e, although the width of each digit may not be the same, the proposed segmentation algorithm can find the boundary of each digit efficiently. This means that the proposed digit segmentation algorithm can locate each digit in ROIs effectively. In addition, to evaluate the performance of the proposed digit segment algorithm, the error rate is defined as the position difference of digit segment between the ground truth and the detected version. The ground truth is obtained by manually locating the ROI in screen images. For 167 digits, the maximum value of errors for the start and end points of digit segments are 5 and 2 pixels, respectively. The average errors are 0.45 and 0.35 pixels for the start and end points of digit segments, respectively. Therefore, the experimental results show that the proposed system can locate these digits in ROIs effectively.

C.2 Digit classification

The Tesseract OCR software created by Google is a well-known tool to recognize digits. Here, the performance of the proposed scheme is compared with that of the Tesseract OCR software. There are 60 samples of each digit for training and there are 674 digits in ROIs for testing. Table 1 shows some wrong results of digit recognition done by the proposed scheme and the Tesseract OCR software. As observed in Table 1, some digits may not be correctly recognized. For case 1 in Table 1, there are one and two errors for the proposed system and the Tesseract OCR software, respectively. In addition, there are 663 and 635 digits correctly recognized by the proposed system and the Tesseract OCR software, respectively. The missing detection and false alarm are 6 and 5 in the proposed system, but the missing detection and false alarm are 27 and 12 in the Tesseract OCR software. This means that the accuracy rates are 98.4% and 94.2% for the proposed system and the Tesseract OCR software, respectively. Furthermore, the accuracy rates of each digit for the proposed system and the Tesseract OCR software are listed in Table 2. The accuracy rates are at least 90.9% for the proposed scheme. As shown in Table 2, the accuracy rates of the digits "8" and "3" are lower than those of the others in both methods. The main reason is that "8" and "3" have little similar shapes in these camera-captured screen images. These experimental results show that the proposed scheme can provide a better performance of digit recognition for camera-captured screen images than the Tesseract OCR software.

6 Discussion

In fact, physicians can use a vision-based device with digit recognition technology to easily recognize the patient identification information (PII) and then find the corresponding medical records. Moreover, the PII can effectively archive patients' skin images into MIS for reducing physicians' working load in dermatology. This means that effectively recognizing PII from camera-captured images is a significant procedure of computer-aided diagnosis in dermatology. To this end, this study aims at developing an efficient vision-based scheme to recognize PII from camera-captured images.

Until now, there are no vision-based systems for recognizing PII captured from camera-captured images. In this paper, a vision-based patient identification recognition system based on image content analysis and support vector machine is proposed for medical information system. The experimental results show that all of the screen images can be correctly identified. Figure 4 illustrates some experimental results of preprocessing, initial ROI cropping, skew correction, candidate region refinement, and

Table 1 Some wrong results of digit recognition (note: the underline means the recognition error and **X** denotes no response)
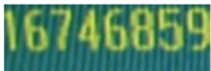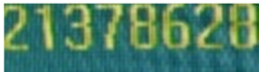
ROIs for testing	Digit recognition	
	Proposed	Tessract
16746859	16746<u>9</u>59	**X** 67 **X** 6859
21378628	2137862<u>3</u>	2137862<u>3</u>
612893	61289<u>4</u>	512393

Table 2 The accuracy rates of each digit for the proposed system and the Tesseract OCR software

	0	1	2	3	4	5	6	7	8	9
Proposed	98.5%	100%	100%	97.2%	98.4%	100%	100%	100%	90.9%	98.4%
Tessract	100%	91.2%	100%	85.4%	96.9%	91.7%	91.3%	100%	88.2%	100%

digit segmentation. These results show that the subparts of the proposed scheme can function well. As for digit recognition, the accuracy rates of digit recognition are 98.4% and 94.2% for the proposed system and the Tesseract OCR software, respectively. The experimental results demonstrate that these extracted features are useful for the digit recognition, and the proposed system is superior to the Tesseract OCR software in terms of the accuracy rate for the digit recognition.

7 Concluding remarks

In this paper, a vision-based patient identification recognition system based on image content analysis and support vector machine is proposed for medical information system, especially in dermatology. The proposed system is composed of three parts: pre-processing, candidate region detection, and digit recognition. To consider the efficiency, the image normalization is performed. The color information is used to identify the camera-captured screen images. In the pre-processing section, a bilateral filter is used to reduce the effect of noise in the captured screen images. The color as well as spatial information is used to initially and roughly locate the candidate region. To reduce the skew effect, a skew correction algorithm based on the Hough transform is developed. A template matching algorithm is used to find the special symbols for locating the region of interest (ROI). For the digit segmentation, digits are segmented in the ROI based on the vertical projection and adaptive thresholding. For the digit recognition, some features are measured for each digit segment and a classifier based on the support vector machine is applied to recognize digits.

To evaluate the performance of the proposed system, a number of images are collected for testing. The experiment results show that the proposed system can effectively not only use the color information to distinguish the captured screen images from the skin images but also detect the ROIs. As for the skew correction, the experimental results show that the proposed skew correction algorithm can effectively decrease the impact of the skew effect on the digit recognition. After the digit segmentation, the average errors are 0.45 and 0.35 pixels for the start and end points of the digit segments, respectively. Furthermore, the accuracy rates of digit recognition are 98.4% and 94.2% for the proposed system and the Tesseract OCR (Optical Character Recognition) software, respectively. The experimental results demonstrate that these features are useful for the digit recognition. In particular, the proposed system outperforms the Tesseract OCR software in terms of the accuracy rate for the digit recognition.

Author details
[1]Department of Computer Science and Information Engineering, National Chin-Yi University of Technology, Taichung, Taiwan. [2]Department of Health Services Administration, China Medical University, Taichung, Taiwan. [3]Department of Computer Science and Information Engineering, Da-Yeh University, Changhua, Taiwan.

References
1. G.-S. Lin, S.-K. Chai, W.-C. Yeh, and Y.-C. Lin, "Suspicious region detection and identification based on intra-/inter-frame analyses and fuzzy classifier for breast magnetic resonance imaging," *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 28, No. 3, pp. 1450007-1- 1450007-26, May 2014.
2. G.-S. Lin, N.-M. Tuan, W.-J. Chen, Detecting region of interest for cadastral images in Taiwan. *Multimed. Tools Appl.* **76**(23), 25369–25389 (2017)
3. M.A. Mohamad, H. Hassan, D. Nasien, H. Haron, A review on feature extraction and feature selection for handwritten character recognition. *Int. J. Adv. Comput. Sci. Appl.* **6**(2), 204–212 (2015)
4. S. Maji and J. Malik, "Fast and accurate digit classification," *Technical Report No. UCB/EECS*-2009-159, November 25, 2009.
5. E. Tuba, M. Tuba, and D. Simian, "Handwritten digit recognition by support vector machine optimized by bat algorithm," *24th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision* (WSCG 2016), pp. 367–376, 2016.
6. M. Sonka, V. Hlavac, and R. Boyle, Image processing, analysis, and machine vision, Thomson, 2008.
7. C. Tomasi, and R. Manduchi, "Bilateral filtering for gray and color images," *Computer Vision*, pp.839-746, Jan. 1998.
8. G.-S. Lin, C.-Y. Chen, C.-T. Kuo, and W.-N. Lie, "A computing framework of adaptive support-window multi-lateral filter for image and depth processing." *IEEE Trans. Broadcast.*, Vol. 60, pp. 452-463, June 2014.
9. Z. Liu, W. Chen, Y. Zou, and C. Hu, "Regions of interest extraction based on HSV color space", Industrial Informatics (INDIN), pp. 481–485, July 2012.
10. C.H. Chou, S.Y. Chu, F. Chang, Estimation of skew angles for scanned documents based on piecewise covering by parallelograms. *Pattern Recogn.* **40**, 443–455 (2007)
11. S. B. Rezaei, A. Sarrafzadeh, and J. Shanbehzadeh, "Skew detection of scanned document images," Proceedings of the International MultiConference of Engineers and Computer Scientists, 2013.
12. A. Alaei, P. Nagabhushan, U. Pal, F. Kimura, An efficient skew estimation technique for scanned documents: an application of piece-wise painting algorithm. *Journal of Pattern Recognition Research* 1, 1–14 (2016)
13. G. Bradski, and A. Kaehler, Learning OpenCV: computer vision with the OpenCV library, O'Reilly Media, Sebastopol, CA, Sep. 2008.
14. Atam P. Dhawan, Medical Image Analysis, John Wiley & Sons, Inc, 2003.
15. C.W. Su, H.Y. Liao, H.R. Tyan, K.C. Fan, L.H. Chen, A motion-tolerant dissolve detection algorithm. *IEEE Trans. Multimedia* **7**, 130–140 (2005)
16. C.L. Huang, B.Y. Liao, A robust scene-change detection method for video segmentation. *IEEE Trans. Circuits and Systems for Video Technology* **11**, 1281–1288 (2001)
17. G. Deng and L. W. Cahill, "An adaptive Gaussian filter for noise reduction and edge detection," *1993 IEEE Conference Record Nuclear Science Symposium and Medical Imaging Conference*, pp. 1615 – 1619, Nov. 1993.
18. G.-S. Lin, M.-K. Chang, Y.-J. Chang, C.-H. Yeh, A gender classification scheme based on multi-region feature extraction and information fusion for unconstrained images. *Multimed. Tools Appl.* **75**(16), 9775–9795 (2016)
19. C.C. Chang, C.J. Lin, LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2**(3), 1–27 (2011)
20. G.S. Lin, M.K. Chang, S.T. Chiu, A feature-based scheme for detecting and classifying video-shot transitions based on spatio-temporal analysis and fuzzy classification. *Int. J. Pattern Recognit. Artif. Intell.* **23**, 1179–1200 (2009)

Publisher's Note