IET Image Processing

Research Article

IET Journals
The Institution of Engineering and Technology

Accurate system for automatic pill recognition using imprint information

ISSN 1751-9659
Received on 8th October 2014
Revised on 29th April 2015
Accepted on 21st May 2015
doi: 10.1049/iet-ipr.2014.1007
www.ietdl.org

Jiye Yu[1], Zhiyuan Chen[1], Sei-ichiro Kamata[2], Jie Yang[1] ✉

[1]*School of Electronic Information and Electrical Engineering, Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai, People's Republic of China*
[2]*Graduate School of Information, Production and System, Waseda University, Kitakyushu, Japan*
✉ *E-mail: jieyang@sjtu.edu.cn*

Abstract: With rapidly advancing of contemporary medicine, it is necessary to help people identify various kinds of pills to prevent the adverse pill events. In this study, a high-accuracy automatic pill recognition system is proposed for accurate and automatic pill recognition. As pill imprint is main distinction between different pills, this system proposes algorithms on both imprint extraction and description parts to make use of imprint information. First, proposed modified stroke width transform is adopted to extract the imprint by detecting coherent strokes of imprint on the pill. Moreover, image segmentation by Loopy belief propagation is also added on printed imprint pills to solve the incoherent and coarse stroke problem. Second, a new descriptor named two-step sampling distance sets is proposed for accurate imprint description and successfully cut down the noise on extracted imprint. This strategy is based on the imprint partition – partitions the imprint on the basis of separated strokes, fragments and noise points. Recognition experiments are applied on extensive databases and result shows 90.46% rank-1 matching accuracy and 97.16% on top five ranks when classifying 12 500 query pill images into 2500 categories.

1 Introduction

As an achievement of contemporary medicine, pills are widely used in the treatment, cure, prevention and diagnosis of disease. However, problems have appeared with the increasing type of pills. Pharmaceutical manufacturers and patients always have no idea to distinguish the unwrapped pills, which possibly even leads to the adverse drug events.

Pill recognition websites are created to help people distinguish different kinds of pills and avoid the adverse drug events. In U.S., Food and Drug Administration (FDA) issued the regulation code 21CFR206 [1] to enforce the unique look for every prescription pill on the market in terms of shape, colour and imprint. These three features become the common marks of one pill, which can be used to distinguish different kinds of pills either by human eyes or by machines. Constructing a stable system to accurately identify pills based on the pills database is what we focus on. By using our system, the automatic recognition of independent pill through its photograph is realisable.

In our system, all shape, colour and imprint information of pills are all utilised in retrieval work. Two operations, imprint extraction and description play a crucial role for successful recognition. However, imprint extracted from pill images by traditional methods always suffers problems such as fragmentary or blurring. Most of unexpected noise and fragments are hard to prevent because of the variance of luminance. Efforts should be done to overcome disturbance in pill retrieval. This paper proposes novel algorithms in both pill extraction and description process to achieve automatic and accurate pill recognition, comparing favourable to the state of the art.

In imprint extraction, we use modified stroke width transform (MSWT) to detect and extract the strokes of imprints. MSWT is comprised of stroke width transform (SWT) [2], switch function and accumulated gradient. SWT is applied on the result of proposed switch function, which leads to a selected rough region including the imprint strokes. Meanwhile, proposed accumulated gradient is applied in gradient calculation instead of the traditional gradient calculation method to cut down noise. In addition, printed imprints (distinguish from the debossed imprints) can further be applied Loopy belief propagation (LBP) [3] based image segmentation algorithm to improve the result extracted by MSWT. The pill surface should be labelled into two parts, the imprint part and the background part. Optimal label choice got by LBP can be the segmentation result.

In pill description, the noise interference also should be suppressed as much as possible. In our work, a descriptor named two-step sampling distance sets (TSDS) is proposed based on distance set [4]. TSDS uses a two-step sampling strategy, where a resampling of feature points according to imprint partition is added to exclude the disturbance of the noise and unwished fragments.

The remainder of the paper is organised as follows: Section 2 reviews the previous work. Section 3 introduces the algorithm of imprint extraction. Section 4 gives the introduction of TSDS. Section 5 presents the construction of the recognition system. Section 6 shows the experimental result. Finally Section 7 concludes the paper and shows the future work.

2 Previous works

In recent years, more and more medical authorities and research institutions pay their attention to the development of medicine recognition system. On the Internet, lots of websites provide pill recognition tools to help people distinguish pills. Several automatic recognition methods have already existed to do pill recognition work using the appearance feature of pills.

2.1 Internet medicine recognition tools

There are several websites on the internet offering the pill recognition tools, such as WebMD Pill Identification Tool [5], RxList Pill Identification Tool [6], Healthline Pill Identifier [7], Pillbox [8] and so on. These recognition tools require people to input pill's shape, colour and imprint or brand. Then tools do the search in the classified database using input information and

output the eligible result. By this way, people can avoid mistaking pills and keep a safe medication.

2.2 Automatic pill recognition methods

Technology that extracts the information from pill images makes batch pill recognition process possible. Many algorithms have been developed to realise this process. Andreas Hartl [9] presents a mobile computer vision system which can take pill size, colour and pill shape into consideration. Young-Beom Lee's algorithm [10] succeeds to use the imprint as key information, in which imprint shape is extracted by means of edge detection method and feature vectors are generated based on edge values using Hu invariant moments [11]. Shape Distribution is introduced in [12] to measure the similarity between 3D shapes. Then Caban [13] and Bukovec [14] make use of this sense and apply it to deal with pill images. Chen introduces the weighted shape context [15] which is based on shape context [16] to describe the pills' imprints. Chen's algorithm has been proved to be effective for dealing with both debossed imprint and printed imprint. However, robustness still needs to be improved when deal with various illumination and noise problems.

A preliminary version of our system is described in [17]. Here are the expansion parts in this work: The algorithm of TSDS has been improved. Two new parts on imprint extraction are proposed: accumulated gradient calculation and image segmentation by LBP. We also update the experimental results.

3 Imprint extraction

As the key information of a pill, imprint should be treated seriously. The first process to collect information from imprint is imprint extraction which aims to split the imprint regions from the pill surface.

Before applying the imprint extraction operations, we do a preprocess on images to enhance the contrast of pills. Adding preprocess step aims to solve the exposure problem. Then we apply the proposed imprint extraction methods.

3.1 Switching function

SWT [2] is a state of art technique for text extraction. However, when dealing with the blurred images, extracted strokes by SWT are always incomplete or full of noise.

In our work, the switch function is proposed to limit the search region for SWT, which helps SWT extract more complete strokes.

The intuitive idea of switch function is trying to find the approximate foreground and background for pills. No matter debossed imprint or printed imprint, almost all imprints parts are darker than other regions on the pill surface. Making use of this property, regions with dark colour on the pill is defined as the foreground, while the background is defined as the distribution of bright colour of the pill. By comparing the original image with the defined foreground and background, a mask image can be generated to show the selected region for imprint.

We suppose there is a bounded function $f(x)$ in one-dimensional (1D) space, and its upper boundary function $F(x)$ and lower boundary function $B(x)$ have the following properties

$$\forall x \in \boldsymbol{R}\ F(x) > f(x), \quad \left|\frac{\mathrm{d}F(x)}{\mathrm{d}x}\right| < \left|\frac{\mathrm{d}f(x)}{\mathrm{d}x}\right|. \tag{1}$$

$$\forall x \in \boldsymbol{R}\ B(x) < f(x), \quad \left|\frac{\mathrm{d}B(x)}{\mathrm{d}x}\right| < \left|\frac{\mathrm{d}f(x)}{\mathrm{d}x}\right|. \tag{2}$$

According to $F(x)$ and $B(x)$, we are able to divide the $f(x)$ into two parts. If $f(x)$ is getting closer to $F(x)$ rather than $B(x)$, we label the value as 1 in mask $\Gamma(x)$. Otherwise, we label the value as 0 in mask $\Gamma(x)$. This process can be written in the following way and illustrated by Fig. 1

$$\Gamma(x) = \begin{cases} 1 & \delta < |B(x) - f(x)| - |F(x) - f(x)| \\ 0 & \text{otherwise} \end{cases}, \quad \text{where } \delta > 0 \tag{3}$$

Here δ is a small value as constraint, which is used to do the fine tuning of $\Gamma(x)$. Since the result mask $\Gamma(x)$ takes the value either 1 or 0, we name it after switch function.

The case in 1D space shows the basic idea of switch function. Switch function can also be used in two-diemsnional spaces to segment the imprint from image I. The upper function $F(x)$ and lower function $B(x)$, respectively, correspond to the foreground function $F(x, y)$ and background function $B(x, y)$ in our case. Mathematical morphology is one optional way to compute foreground and background functions. Considering that each stroke on imprint usually consists of two parts: bright part and dark part

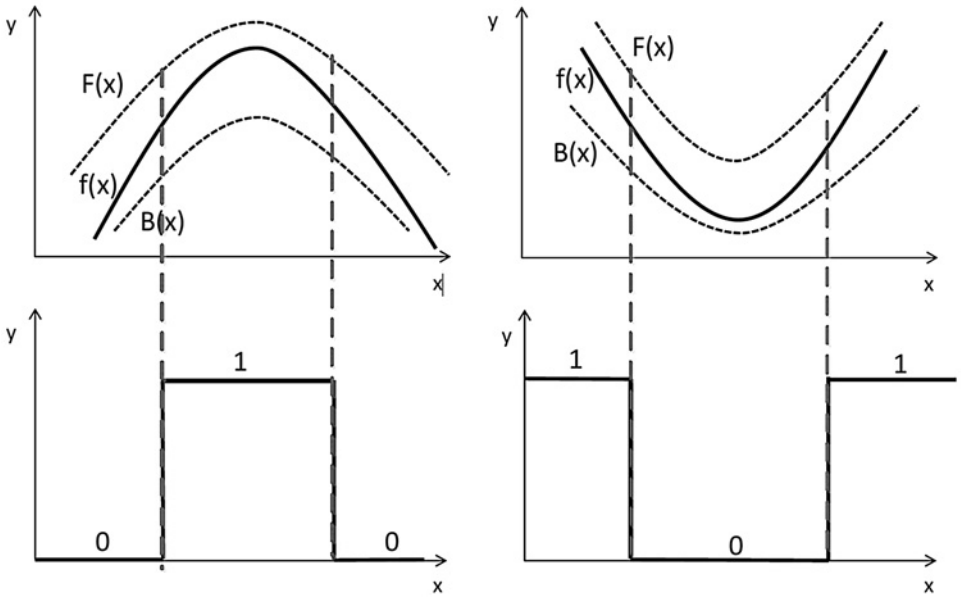

Fig. 1 *In 1D case, the switch functions on the second row are the results by comparing the f(x) with the distance to its corresponding foreground function F(x) and background function B(x)*

IET Image Process., 2015, Vol. 9, Iss. 12, pp. 1039–1047

(especially for debossed imprints), we make use of switch function to get rough imprint region for both two parts. For bright part, the morphological erosion of $I(x, y)$ is used to represent the foreground function $F_b(x, y)$ and the morphological closing of $I(x, y)$ is used to represent the background function $B_b(x, y)$. For dark part, the morphological dilation of $I(x, y)$ is used to represent the foreground function $F_d(x, y)$ and the morphological opening of $I(x, y)$ is used to represent the background function $B_d(x, y)$. Here the subscript 'b' and 'd' stand for bright part and dark part, respectively. The selected region to the given imprint, $\Gamma_b(x, y)$ of bright part and $\Gamma_d(x, y)$ of dark part can be shown as

$$\Gamma_i(x, y) = \begin{cases} 1 & \delta < \left|B_i(x, y) - I_i(x, y)\right| - \left|F_i(x, y) - I_i(x, y)\right| \\ 0 & \text{otherwise} \end{cases}, \quad \text{for } \forall(x, y) \in I. \tag{4}$$

where suffix i can be 'b' (the bright part) or 'd' (the dark part). δ is still a small value as constraint to prevent redundant selected region appearing. After repeated experiments, we decide $\delta \in [5, 10]$ in our case, and the δ value of bright part and dark part can vary from each other.

The combination of $\Gamma_b(x, y)$ and $\Gamma_d(x, y)$ by operation OR generates the selected region $\Gamma(x, y)$ for the imprint on given pill image

$$\Gamma(x, y) = \Gamma_b(x, y) \vee \Gamma_d(x, y) \text{ for } \forall(x, y) \in I. \tag{5}$$

This selected region $\Gamma(x, y)$ is the most possible region where imprint can occur.

Switch function ensures that the computed edge map by edge detector can fully represent the boundary of imprints. Applying SWT in selected region on pill image can greatly improve the extracted imprint and reduce noises.

3.2 Accumulated gradient calculation

Gradient direction at pixel p, $\boldsymbol{d}_p$, is used to find p's corresponding point on the stroke in vertical direction in SWT. The traditional gradient direction is usually expressed as

$$\boldsymbol{d}_p = (gX(x, y), gY(x, y)). \tag{6}$$

Here $gX(x, y)$ and $gY(x, y)$ are magnitudes at (x, y) computed by means of Sobel, Prewitt or other edge detectors. However, in pill cases, noises or biases originally existing in original image and generated in image resizing may make the calculated gradient direction unreliable for particular pixel. Inspired by the idea in bilateral filter, we try to use a similar accumulation mechanism to computer the gradient direction. The basic idea of accumulated gradient calculation is to get a reliable gradient direction by accumulating the gradient direction in a small neighbour region. By counting the gradient in a local neighbourhood, the calculated gradients tend to be smoothed, which means there will be less sudden change between adjacent pixels. This kind of smoothness in gradients of adjacent pixels ensures that their corresponding points on the opposite of the stroke are also not far away.

When doing the accumulation for one particular pixel, we consider both the spatial distance and angular difference between the centre pixel and its adjacent pixels. For centre pixel $p_0 = (x_0, y_0)$ with gradient $\boldsymbol{d}_{p_0} = (gX(x_0, y_0), gY(x_0, y_0))$ and its adjacent points $p_1 = (x_0 + \Delta x, y_0 + \Delta y) = p_0 + \Delta$, where $\Delta = (\Delta x, \Delta y)$ with gradient $d_{p_1} = (gX(x_0 + \Delta x, y_0 + \Delta y), gY(x_0 + \Delta x, y_0 + \Delta y))$, their spatial distance is measured by Gaussian function with Euclidean distance between its arguments

$$d_s(p_0, p_1) = \mathrm{e}^{(1/2)\left((\|p_0 - p_1\|_2)/\sigma_d\right)^2}. \tag{7}$$

Since we only count pixels with similar gradient direction in accumulation, pixels with a more than 90° angular difference are all set to 0. This process can be written as

$$d_a(p_0, p_1) = \begin{cases} 1 & \boldsymbol{d}_{p_0} \cdot \boldsymbol{d}_{p_1} > 0 \\ 0 & \boldsymbol{d}_{p_0} \cdot \boldsymbol{d}_{p_1} \le 0. \end{cases} \tag{8}$$

The final accumulation functions are shown as follows

$$GX(x_0, y_0) = \sum_{\Delta x=-R}^{R} \sum_{\Delta y=-R}^{R} d_s(p_0, p_1) d_a(p_0, p_1) gX(p_1). \tag{9}$$

$$GY(x_0, y_0) = \sum_{\Delta x=-R}^{R} \sum_{\Delta y=-R}^{R} d_s(p_0, p_1) d_a(p_0, p_1) gY(p_1). \tag{10}$$

Here R is the pre-determined accumulation region size. Then gradient direction can be calculated as in (6) by substituting $gX(x_0, y_0)$ and $gY(x_0, y_0)$ with $GX(x_0, y_0)$ and $GY(x_0, y_0)$.

Combining switching function and accumulated gradient, MSWT can get a better result than SWT. Experimental results on this part are shown in Section 6.1.

3.3 Image segmentation by LBP

Interrupted coarse strokes and defective fragments occasionally appear on the extracted imprints by MSWT. As the imprint extraction process can be regarded as image segmentation, we can use LBP algorithm to solve the markov random field (MRF) model [18] to partition the whole pill image into two segments (background segment and imprint segment) and get a smoother and more complete imprint.

LBP algorithm over pairwise-connected MRF model has become widely used for low-level vision problems. As LBP is a local message passing algorithm for performing inference on graphical model. We first make the MRF model for our imprint extraction case, and then use LBP algorithm to make a solution.

Using the result extracted by SWT, we label the imprint part as 1 and label the background part as 0. This can be regarded as the pre-segmentation. Based on the pre-segmentation, obtaining a complete imprint is equal to getting an optimal label choice which makes the energy function minimum

$$E(f) = \sum_{(p,q)\in\varpi} V(f_p, f_q) + \sum_{p} D(f_p). \tag{11}$$

Here f represents the labelling choice. $V(f_p, f_q)$ is the non-smooth cost. $D(f_p)$ is the data cost. ϖ is the connected edge in four-connected neighbour system.

We apply Gaussian mixed model on pill images to do the calculation of data cost. Two Gaussian peaks can be got on the printed imprint pill images, which represent the imprint part and the background part, respectively. With the help of pre-segmentation, we can calculate each peak's parameters (μ_k, σ_k, $k = \{1, 2\}$) in advance. The data cost function is

$$D_p(f_p) = \ln(2\pi\sigma_k^2)^{(1/2)} + \frac{(x_p - \mu_k)^2}{2\sigma_k^2}. \tag{12}$$

Here $f_p \in L$, $L = \{0, 1\}$. L is the label space.

The smoothness cost function of Potts model is

$$V(f_p, f_q) = \beta \times \delta(f_p, f_q). \tag{13}$$

$$\delta(f_p, f_q) = \begin{cases} 0 & f_p = f_q \\ 1 & f_p \ne f_q. \end{cases} \tag{14}$$

Then we apply the Loopy belief propagation algorithm [3, 19] to

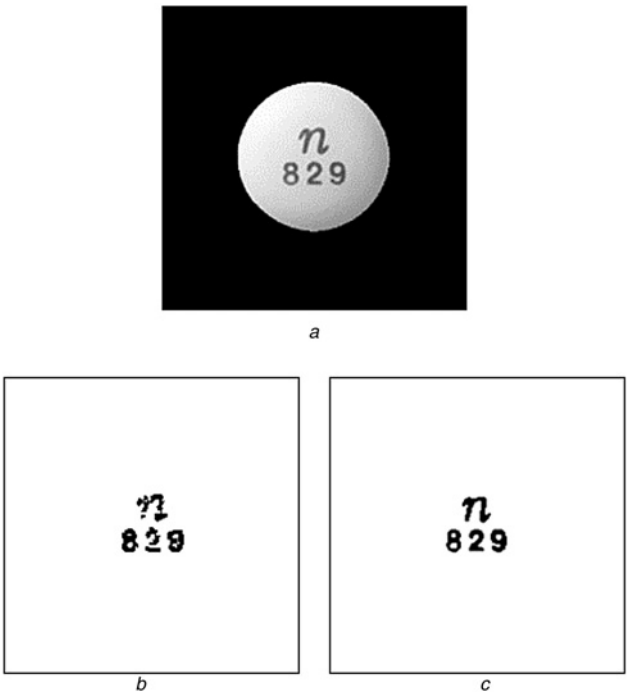

Fig. 2 *An example shows the result of LBP image segmentation*
a Original pill image
b Its imprint extracted by MSWT
c Repaired imprint by LBP image segmentation

obtain the solution

$$m_{pq}^{t}(f_q) = \min_{f_p}\left(V(f_p, f_q) + D_p(f_p) + \sum_{s\in N(p)\setminus q} m_{sp}^{t-1}(f_p)\right), \quad (15)$$

$$b_q(f_q) = D_q(f_q) + \sum_{p\in N(q)} m_{pq}^{T}(f_q). \quad (16)$$

The label f_q^* that minimises $b_q(f_q)$ is exactly the optimal label for point q.

We apply this process only on pixels in selected region got by switch function. A smooth imprint will be got after this process. Fig. 2 shows an example. We can find that interrupted coarse strokes and defective fragments are repaired.

We should note that to the debossed imprint pill images, because of the illumination problem, it is hard to distinguish imprint from the shadow appearing on the pill surface. In fact, the debossed imprint is one kind of the shadows in 2D vision. To the debossed pill images, LBP algorithm may be not effective.

4 Imprint description using two-step sampling distance sets

Unlike shapes in the usual sense, imprint images always consist of several regions. However, it still can be treated as a shape and described by kinds of shape descriptors. Descriptor TSDS is proposed for imprint images, but it can also be used on other shape images with the similar characteristic such as trademarks and road signs.

Distance set introduced by Grigorescu and Petkov [4] is a kind of rich local descriptor. As a local descriptor, distance set describes the distance between a given point and its K nearest neighbouring points on the shape's contour. The utilisation of relative distance helps it be invariant to rotation. After applied it to every sampling point for a pill image, set of distance sets can be got by assembling these distance sets. In summary, distance set describes a local arrangement of points around the specified point, and set of distance sets describes the global spatial arrangement of the whole image.

4.1 Uniform sampling

Contour contains discriminative information of one shape. Since there are too many points on the contour, before applying a descriptor, shape contour always needs sampling.

Generally, uniform sampling is commonly used because it can preserve the primitive structure of shape well. Sampling points should be uniformly distributed across the contour of the whole imprint. As shown in Fig. 3, black points are the sampled points on the contour of the whole imprint. A certain interval is set between two sampled points.

4.2 Imprint partition

Different from single closed curve shape, the imprint shapes are more complex and irregular. As an imprint usually consists of some letters, symbols and other kinds of marks, it can always be divided into several separated regions. This partition step can be named after imprint partition. Here regions are defined as independent strokes on the imprint. If noise points exist, they can also be regarded as regions.

As a result, whether two points are in the same connected stroke is regarded as the criterion to decide whether two points belong to one region or not. As the imprint shown in Fig. 4*b*, there are six major regions composing five letters ('*N*', '*S*', '1', '1', '6'; here '*S*' is constituted by two regions because there is a breakage in the middle of the stroke) and several noise regions. Fig. 4*c* shows the regions in different colours (totally there are 13 regions in this case).

As a region is constituted by points, its region size can be evaluated by the number of points on its contour. Generally, these regions are always not even sized, and noise regions are relatively much smaller.

If there are N points on the contour of imprint, for the kth region $R_k = \{p_1, p_2, \ldots, p_{\#R_k}\}$, we define its region size as $\#R_k$, where $\#R_k$ is the numbers of points in kth region. We define the size ratio of kth region as the ratio between size of kth region ($\#R_k$) and size of whole imprint contour (N): ($\#R_k/N$).

4.3 Two-step sampling strategy and two-step sampling distance sets

Based on the imprint partition, we propose the two-step sampling strategy which aims to remove the noise from the imprint description.

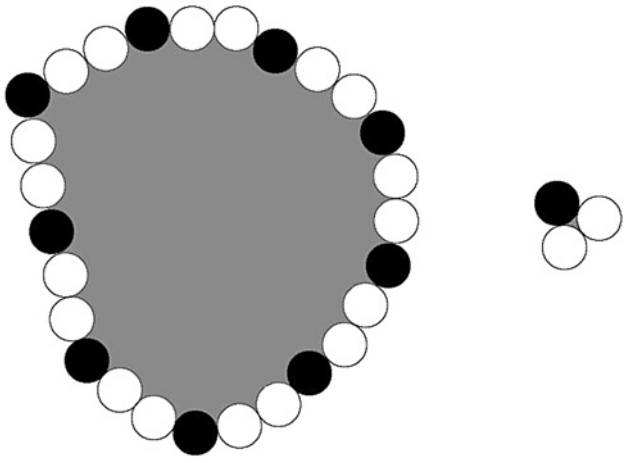

Fig. 3 *Example of uniform sampling*
Black points are the sampled points on the contour of this shape

IET Image Process., 2015, Vol. 9, Iss. 12, pp. 1039–1047

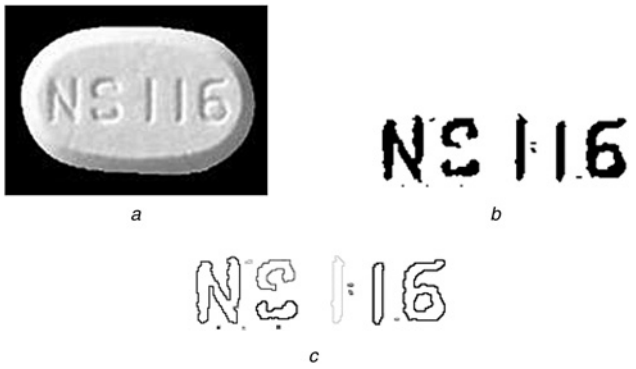

Fig. 4 *An example of pill, imprint, and regions*

a Original pill image
b Its extracted imprint
c Regions in different colours

The uniform sampling can keep the general structure of the origin contour of imprint, as well as regions' original size ratio. However, noise is also reserved after the uniform sampling. To eliminate noise, we propose the two-step sampling strategy.

The basic idea of second step sampling is

(1) The smaller a region's size ratio is, the less point it should be sampled.
(2) A region relatively larger than others retains more points.

By this way, noise regions whose region size is always small will be cut down.

Following is the detailed procedure of two-step sampling strategy. First, we need to apply the uniform sampling on the imprint contour and get N_1 sampling points. For kth region, its region size is $\#R_{1k}$ under the uniform sampling. We can obtain

$$N_1 = \sum_{k=1}^{T} \#R_{1k},\ k \in [1, T]. \tag{17}$$

Here T is the number of regions. We also need to decide the number of second step sampling points $N_2 (N_2 < N_1)$ in advance. Then based on the basic idea of second step sampling, the size of kth region on second step sampling is decided as

$$\#R_{2k} = \#R_{1k} - \left(\frac{\#R_{1k}}{N_1}\right)^{\gamma} \cdot \frac{1}{r} \cdot (N_1 - N_2), \quad k \in [1, T]. \tag{18}$$

here

$$r = \sum_{k=1}^{T} \left(\frac{\#R_{1k}}{N_1}\right)^{\gamma}$$

is the curvature parameter which is used to do the normalisation, aiming to ensure

$$N_2 = \sum_{k=1}^{T} \#R_{2k}, \quad k \in [1, T].$$

$\gamma \in (0, 1]$. The smaller γ is, the more complete elimination of small regions will be done. The choice of γ is decided by experiments shown in Section 6.2. It should be noted that when γ is close to 0, sometimes we can get an exceptional case that $\#R_{2k} < 0$ happens. If that case happens, we can set $\#R_{2k} = 0$, and other regions apportion the lacking points.

After that, a new uniform sampling is applied on each kth region's contour with the determined $\#R_{2k}$. Fig. 5*a* shows an example.

Fig. 5*b* shows the second step sampling result with the setting of parameters: $N_1 = 100$, $N_2 = 90$, $\gamma = 0.1$. We let $N_1 - N_2$ be slightly larger than the estimated number of sampled noise points. Then in the figure, we can find that the noise points are cut down completely while the number-fixed feature points are in a quite uniform distribution.

From [4], we can get the expression of local descriptor distance set of point p to its K nearest neighbouring points within the sampling points of shape $S = \{R_1, R_2, \ldots, R_T\}$ like this

$$\mathrm{DS}_{S,K}(p) = \{d_1(p), d_2(p), \ldots, d_i(p), \ldots, d_K(p)\}. \tag{19}$$

Here $d_i(p)$ is the distance between point p and its ith nearest neighbour from shape S, $1 \le i \le K$. Fig. 5*c* shows an example. If $K = 3$, the black left-hand side bottom point p should find three nearest neighbouring points (connected by blue lines). Then calculate the distances between them and record the distances as a descriptor.

After using the two-step sampling strategy, TSDS can be expressed as

$$\mathrm{RDS}(R_{2k}) = \{\mathrm{DS}_{S,K}(p_1), \mathrm{DS}_{S,K}(p_2), \ldots, \mathrm{DS}_{S,K}(p_{\#R_{2k}})\}. \tag{20}$$

$$\mathrm{TSDS}_K(S) = \{\mathrm{RDS}(R_{21}), \mathrm{RDS}(R_{22}), \ldots, \mathrm{RDS}(R_{2k})\}. \tag{21}$$

where R_{2k} is the kth region after second step sampling. Distance measure between TSDS descriptors is the distinction between two shapes with their respective N_2 feature points. Let $\psi: S_1 \mapsto S_2$ be a point-to-point mapping from S_1 to S_2 and let Ψ be the set of all such mappings. Then a cost of the mapping $\psi \in \Psi$ is defined as follows

$$C_K^{(\psi)}(S_1, S_2) = \frac{1}{N_2}\left(\sum_{i=1}^{N_2} D_{S_1,K;S_2,K}(p_i, \psi(p_i))\right) \tag{22}$$

where $D_{S_1,K;S_2,K}(p, q)$ is a dissimilarity between two distance sets $\mathrm{DS}_{S_1,K}(p)$ and $\mathrm{DS}_{S_2,K}(q)$. Dissimilarity between $\mathrm{TSDS}_K(S_1)$ and

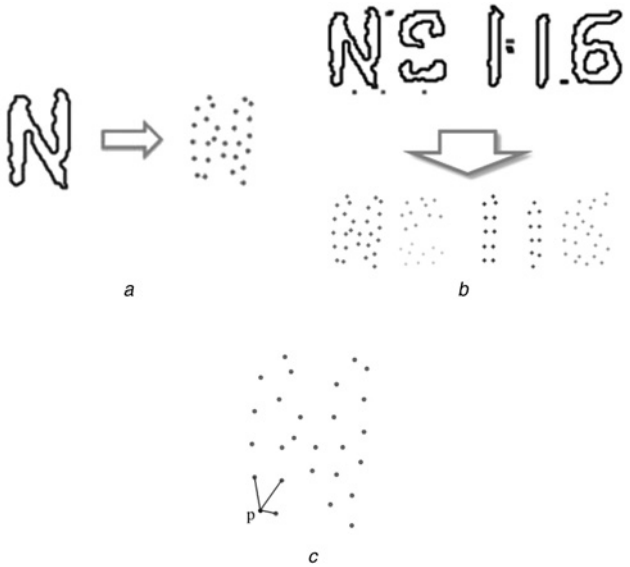

Fig. 5 *Second step sampling and example of nearest neighbour points*

a Uniform sampling for each region with the decided number of sampling points
b Result of second step sampling
c Point p and its three nearest neighbouring points

$\mathrm{TSDS}_K(S_2)$ is defined as

$$\varepsilon_K(S_1, S_2) = \min\{C_K^{(\psi)}(S_1, S_2) | \psi \in \Psi\}. \quad (23)$$

TSDS descriptor is applied to describe the imprint image.

Computation of the dissimilarity between two TSDS descriptors can be reformulated in terms of minimum weight assignment problem in bipartite graph by Kuhn–Munkras algorithm [20]. It can be solved efficiently in $O(v \times (e + v \log(v)))$, ($v$ and e are the number of vertices and edges of the associated graphs, respectively), which is still a little bit time-consuming when cardinality of set of distance sets is large.

5 Construction of recognition system

Our recognition system contains several parts, including the extraction and representation of different features. A system flowchart is shown in Fig. 6. In our recognition system, besides imprint feature, pill shape and colour features should also be taken into account to improve efficiency. These two features are used to realise a rough selection of pill categories, which enables to discard unrelated categories with completely different shape and colour. Only categories passed the first step selection are further applied with imprint based recognition process. The combination of three kinds of features enables to construct a fast and accurate recognition system.

5.1 Pill shape feature

In our method, we use a vector $\boldsymbol{V}_i$ to represent the pill shape feature of pill i

$$\boldsymbol{V}_i = (c_{i,1}, c_{i,2}, \ldots, c_{i,M}). \quad (24)$$

We need to uniformly sample M points on the pill external boundary, and $c_{i,j}$ in (24) is the Euclidean distance between the centre of pill and the jth point. Fig. 7 shows an example.

Cross-correlation $r(V_a, V_b)$ is used as a standard to evaluate the similarity of two pills' shape features V_a and V_b. The definition of cross-correlation $r(V_a, V_b)$ is

$$r(V_a, V_b) = \max\left(\sum_{j=1}^{M} V_{a,j} V_{b,(j+k)M}\right), \quad (0 \le k < M). \quad (25)$$

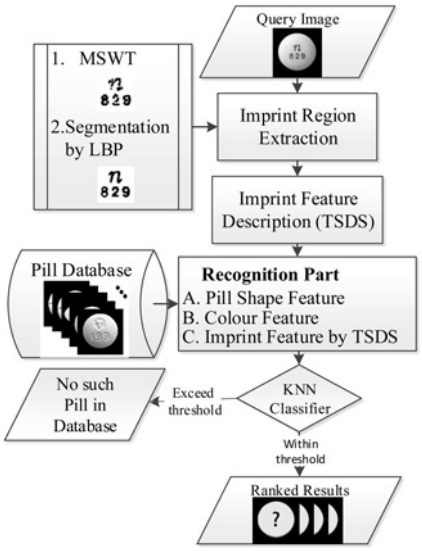

Fig. 6 *Flowchart of the recognition system*

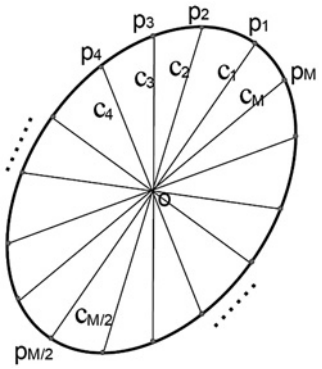

Fig. 7 *Uniformly sample M points on the boundary*

Construct the shape feature by computing the distance between centre O and each point

Here the subscript $(j + k)$ mod M introduces cyclic shifting property. The maximum value is obtained at the position when two shape are aligned.

5.2 Colour feature

When dealing with the colour feature, to eliminate the disturbance of luminance, we convert the pill images into HSV colour model. V channel can be removed because it has no contribution to the colour information. Another reason of using HSV model not RGB model is that colour in HSV model is more similar to the way human eye perceiving colour.

We build colour histogram to construct the colour feature. In colour feature comparing, the metric for histogram matching uses intersection as the following formula [21, 22] to achieve quick comparison

$$d_{\text{intersection}}(H_1, H_2) = \sum_i \min(H_1(i), H_2(i)). \quad (26)$$

Here H_1 and H_2 are two histograms built from different pill images.

6 Experimental results

We have built an image-capturing device to construct the pill database and evaluate the performance of our pill recognition system. The image-capturing device contains a white light source and a dark fluffy plane. A parclose is around the plane to reduce the influence caused by ambient light. Our database acquired by this device consists of 2500 different pill categories, and each category provides at least one image for corresponding pill. Fig. 8 shows samples of pill images in the database. To verify the robustness on rotation and illumination, query images are generated by randomly rotating the images from 0 to 359 degrees, as well as randomly changing the brightness from −30% to 0. We prepare five query images for each category so that totally there are 12 500 query images in the database. As for the resolution, all the pill images are normalised into the size of 200 × 200 pixels.

6.1 Comparison on imprint extraction

The first experiment aims to compare MSWT and SWT on imprint extraction. Fig. 9 shows three pill images and their imprints extracted by two methods. We can find that the switch function succeeds to select an exact area where the imprint stays. Noises can be cut down by our proposed MSWT comparing with SWT.

6.2 Parameter estimation for curvature parameter γ

To evaluate the various selections of parameter γ, we pick out 500 pills categories whose imprints are similar with their query images

IET Image Process., 2015, Vol. 9, Iss. 12, pp. 1039–1047

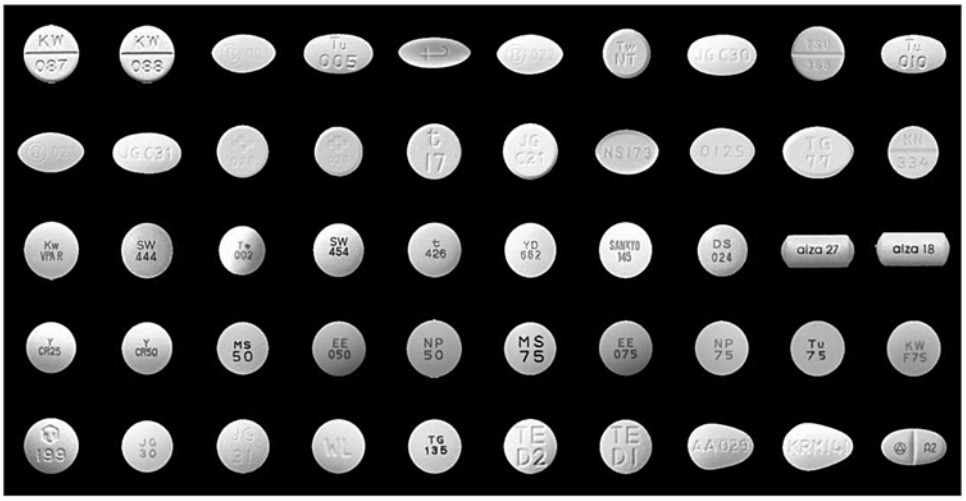

Fig. 8 *Examples of the pill images in database*

from the database to do the retrieval experiments. Equation (18) sets a standard to decide the region size of the second step sampling. Different selection of curvature parameter γ will affect the final retrieval result. $\gamma = 1$ means that the second sampling makes no change on the first uniform sampling. The smaller γ becomes, the more complete elimination of small regions will be done. In other words, the smaller γ becomes, the more noises will be cut down.

Table 1 shows the recognition results by TSDS under various selections of curvature parameter γ. We select five most similar categories for each query pill by KNN classifier. The most similar category is rank-1st, and each column of candidates (accuracy) means the probability of the correct category appears in one of the first i ranks. We can find that a small γ leads to a better result than uniform sampling ($\gamma = 1$), which also means noise is successfully cut down by two sampling strategy. When γ ranges between 0.08 and 0.3, we can obtain a relatively higher accuracy. A small γ leads to a better noise reduction effect. However when γ is less than 0.08, some smaller imprint regions will be treated as noise. As a result, the imprint becomes incomplete and accuracy starts to get worse.

6.3 Comparison with other recognition algorithms

In this section, we show the results of comprehensive experiments using full database which contains whole 2500 categories and 12 500 query images. This database is challenging because pills in the database have similar scale and shape and same series of pills even have the similar imprints. In the experiment, our descriptor is compared with other existed ones under different evaluation criteria. All descriptors use the imprints extracted by our proposed imprint extraction method.

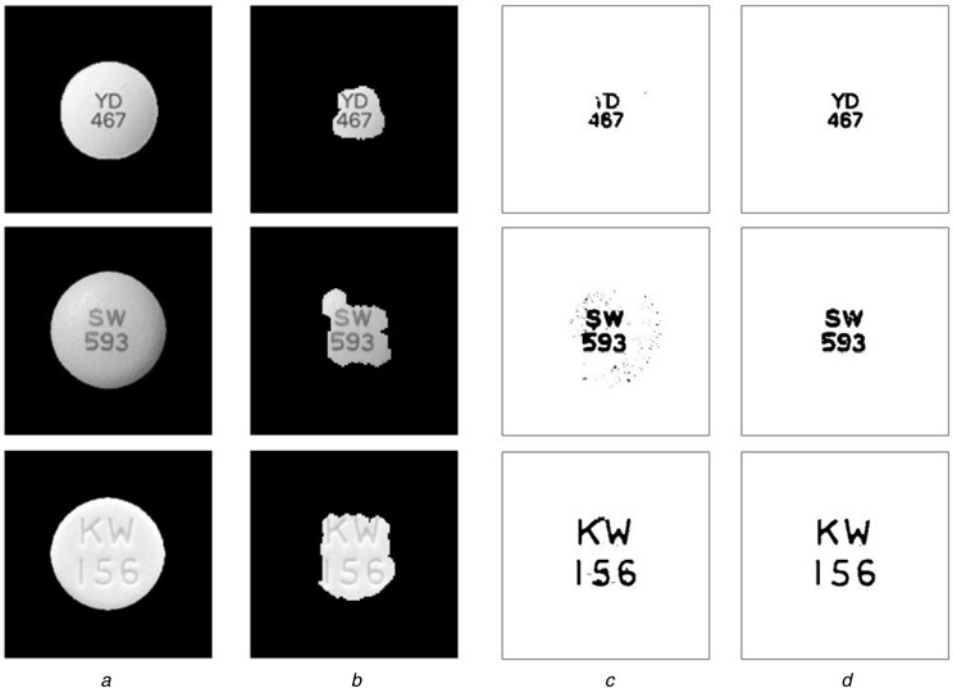

Fig. 9 *Three pill images and their imprints extracted by two methods*

a Original pill images
b Selected regions by using proposed switch function
c Extracted imprints by SWT
d Extracted imprints by MSWT which containing switch function and accumulated gradient

Table 1 Recognition results by TSDS under various γ

γ	Candidates (accuracy)				
	Rank-1st, %	2nd, %	3rd, %	4th, %	5th, %
1	93.16	96.92	97.72	98.12	98.76
0.7	93.68	97.00	97.80	98.28	98.92
0.5	93.72	97.04	97.80	98.36	98.96
0.3	93.88	97.32	98.08	98.44	99.00
0.1	94.00	97.48	98.20	98.48	99.08
0.08	93.92	97.44	98.16	98.44	99.04
0.05	93.76	97.36	98.00	98.32	98.92

Table 2 Comparison of retrieval accuracies at top five ranks

Method	Candidates rank (accuracy)				
	Rank-1st, %	2nd, %	3rd, %	4th, %	5th, %
shape distribution	34.31	45.10	51.58	57.02	60.92
Hu moments	37.55	51.68	60.28	66.51	71.13
weighted shape context [16]	83.25	89.66	92.23	93.43	94.23
proposed method	90.46	94.55	95.97	96.70	97.16

Table 2 shows the comparison result of retrieval accuracy between the proposed TSDS and other descriptors. We select five most similar categories for each query pill using KNN classifier. Parameter of TSDS is set as $N_1 = 100$, $K = 50$ (nearest neighbouring points), $N_2 = 85$ and $\gamma = 0.3$. The average processing speed of our system is around 2.12 s per query pill under the condition of Core(TM)2 Duo CPU E7200, 2.53 GHz.

We also calculate the false rejection rate (FRR) and false acceptance rate (FAR) to evaluate the algorithms. Fig. 10 shows the ROC curves obtained by varying thresholds on imprint descriptors and recording the points (FAR, FRR). As FAR and FRR cannot reach a quite minimal value under the same threshold, the ROC curve shows a visual characterisation of the trade-off between the FAR and the FRR. From the ROC curves we can find that our proposed algorithm performs better than existing methods.

Comparing with [17], imprint extraction part in this work has been improved by accumulated gradient Calculation and LBP image segmentation step. In addition, two-step sampling strategy in TSDS has also been modified. Instead of directly deleting $(N_1 - N_2)$ points on the first step sampling result to make the second step sampling, we do the uniform sampling on each region to compose the second step sampling (details are shown in Fig. 5) in this work. This improvement enables us to preserve the primitive structure of imprint. As a result, the precision of experiment rises from 86.31 to 90.46% on rank-1 and 93.76 to 97.16% on top five ranks.

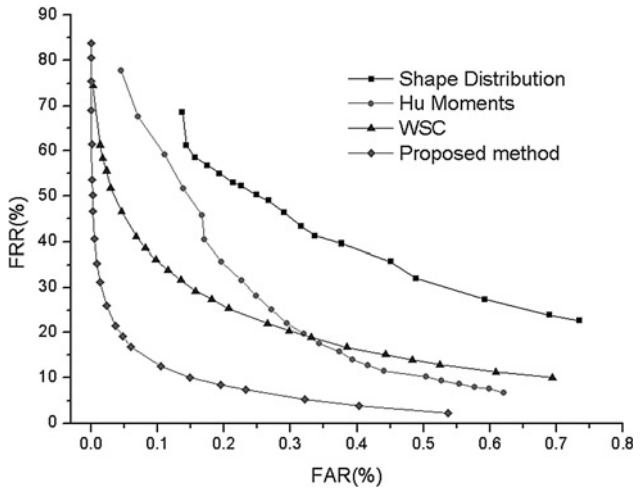

Fig. 10 *ROC curves for descriptors. Each point in the plot shows a pair of FAR and FRR under a selected threshold*

7 Conclusion

In this paper, a pill recognition system is proposed, which can classify the pills with high accuracy. This system combines the efforts of pill's shape, colour and imprint features and imprint plays the most decisive role. In the process of imprint extraction, we propose the MSWT to detect the imprint strokes. The switch function and accumulated gradient calculation can effectively eliminate the noise in result and maintains the continuity of each stroke. The using of LBP algorithm on image segmentation improves the result. It makes the imprint smoother and more complete. In imprint description part, the distance sets is modified by two-step sampling strategy to build the TSDS descriptor which can reduce the disturbance of noise and unwished fragments on imprint images caused by variation of luminance and exposure. Even in the condition of relatively low resolution, our system still achieves 97.16% accuracy among top five matches when testing with 12 500 query images. This pill recognition system can be applied to the mass production in pharmaceutical factories as doing quality inspection. It can also be used to confiscate the illicit drugs.

As for the future work, we will focus on accelerating the algorithms. In addition, we also intend to continue the research of imprint extraction, especially when the pill images are in bad conditions.

8 Acknowledgment

This work is partially funded by NSFC key project No. 61133009, 'Techniques for Complicated Scene Modelling and Super-high Resolution Rendering' and JSPS KAKENHI Grant Number 24500216.

9 References

1 'CFR- code of federal regulations title 21', U.S. Food and Drug Adminstration, available at http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=211, accessed January 2015
2 Epshtein, B., Ofek, E., Wexler, Y.: 'Detecting text in natural scenes with stroke width transform'. Int. Conf. Computer Vision and Pattern Recognition, IEEE, 2010, pp. 2963–2970
3 Felzenszwalb Pedro, F., Huttenlocher, D.P.: 'Efficient belief propagation for early vision', *Int. J. Comput. Vis.*, 2010, **70**, (1), pp. 41–54
4 Grigorescu, C., Petkov, N.: 'Distance sets for shape filters and shape recognition', *IEEE Trans. Image Process.*, 2003, **12**, (10), pp. 1274–1286
5 'WebMD pill identification tool', available at http://www.webmd.com/pill-identification/, accessed January 2015
6 'RxList pill identification tool', available at http://www.rxlist.com/pill-identification-tool/article.htm, accessed January 2015
7 'Healthline pill identifer', available at http://www.healthline.com/pill-identifier, accessed January 2015
8 'Pillbox', National Library of Medicine, National Institutes of Health, United States, available at http://pillbox.nlm.nih.gov/, accessed January 2015
9 Hartl, A.: 'Computer-vision based pharmaceutical pill recognition on mobile phones'. Proc. 14th Central European Seminar on Computer Graphics, May 2010, p. 51
10 Lee, Y.B., Park, U., Jain, A.K.: 'Pill-ID: Matching and retrieval of drug pill imprint images'. Proc. 20th Int. Conf. Pattern Recognition, IEEE, August 2010, pp. 2632–2635
11 Hu, M.K.: 'Visual pattern recognition by moment invariants', *IRE Trans. Inf. Theory*, 1962, **8**, (2), pp. 179–187
12 Osada, R., Funkhouser, T., Chazelle, B., Dobkin, D.: 'Shape distributions', *ACM Trans. Graphics*, 2002, **21**, (4), pp. 807–832
13 Caban, J.J., Rosebrock, A., Yoo, T.S.: 'Automatic identification of prescription drugs using shape distribution models'. 19th Int. Conf. Image Processing, IEEE, 2012, pp. 1005–1008
14 Bukovec, M., Špiclin, Ž., Pernuš, F., Likar, B.: 'Automated visual inspection of imprinted pharmaceutical tablets', *Meas. Sci. Technol.*, 2007, **18**, (9), p. 2921
15 Chen, Z., Kamata, S.: 'A new accurate pill recognition system using imprint information'. Sixth Int. Conf. Machine Vision, London, UK, December 2013, p. 906711
16 Belongie, S., Malik, J., Puzicha, J.: 'Shape matching and object recognition using shape contexts', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2002, **24**, (4), pp. 509–522

17 Yu, J., Chen, Z., Kamata, S.: 'Pill recognition using imprint information by two-step sampling distance sets'. In 22nd Int. Conf. Pattern Recognition, IEEE, Stockholm, Sweden, August 2014
18 Li, S.Z.: 'Markov random field models in computer vision'. Computer Vision' (Springer Berlin Heidelberg, 1994), pp. 361–370
19 'Loopy belief propagation, Markov Random Field, stereo vision'. Nghia Ho, available at http://nghiaho.com/?page_id=1366, accessed January 2015
20 Kuhn, H.W.: 'The Hungarian method for the assignment problem', *Nav. Res. Logist. Q.*, 1955, **2**, (1–2), pp. 83–97
21 Smeulders, A.W., Worring, M., Santini, S., *et al*.: 'Content-based image retrieval at the end of the early years', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2000, **22**, (12), pp. 1349–1380
22 Niblack, C.W., Barber, R., Equitz, W., *et al*.: 'QBIC project: querying images by content, using colour, texture, and shape'. IS&T/SPIE's Symp. on Electronic Imaging: Science and Technology, 1993, pp. 173–187