A New Accurate Pill Recognition System Using Imprint Information

Zhiyuan Chen and Sei-ichiro KAMATA
Graduate School of Information Production and System
Waseda University
Kitakyushu, Japan
zychen_ips@moegi.waseda.jp, kam@waseda.jp

ABSTRACT

Great achievements in modern medicine benefit human beings. Also, it has brought about an explosive growth of pharmaceuticals that current in the market. In daily life, pharmaceuticals sometimes confuse people when they are found unlabeled. In this paper, we propose an automatic pill recognition technique to solve this problem. It functions mainly based on the imprint feature of the pills, which is extracted by proposed MSWT (modified stroke width transform) and described by WSC (weighted shape context). Experiments show that our proposed pill recognition method can reach an accurate rate up to 92.03% within top 5 ranks when trying to classify more than 10 thousand query pill images into around 2000 categories.

Keywords: pill recognition, Modified Stroke Width Transform (MSWT), Weighted Shape Context (WSC), feature extraction, image retrieval.

1. INTRODUCTION

The significant development of modern medicine greatly benefits human beings. Pharmacy industry provides numerous kinds of pharmaceuticals to cure different diseases. However, it also raises problems. For people without being professional trained, it is impossible to make out the pill when found unlabeled in daily life. Hence, to find an effective pill recognition method is becoming an urgent topic. Appearance features of pills are believed to be helpful. As in U.S., regulation code 21CFR206 [1] issued by Food and Drug Administration (FDA) enforces the unique look for every prescription pill sold in the market in terms of size, shape, color, and imprint. If these visual appearance features can be correctly extracted, pills can be accurately identified with the help of a pill database got beforehand.

In this study, the main issue is to construct a recognition mechanism making full use of the imprint information. Two steps are designed. First, a step named imprint extraction is applied on the input images. Modified Stroke Width Transform (MSWT) is used to draw a clear imprint figure from the input. The extracted imprint then goes into imprint description step, where imprint features are coded into a set of histograms under our proposed Weighted Shape Context (WSC). This configuration of our method has been proved to be effective for dealing with both debossed imprint and printed imprint.

Figure 1. Two typical types of pill imprint. (a) is debossed imprint. (b) is printed imprint.

The rest of the paper is organized as follows. In the next section, we review some existing works. Section 3 gives a detail introduction of our imprint extraction and imprint description steps. In section 4, the whole pill recognition system is explained briefly. Our datasets and experimental results are presented in section 5 and we conclude in section 6.

2. PERVIOUS WORK

According to working manners, existing pill recognition systems fall into two categories: manual input method and automatic recognition method.

Sixth International Conference on Machine Vision (ICMV 2013), edited by Antanas Verikas, Branislav Vuksanovic, Jianhong Zhou, Proc. of SPIE Vol. 9067, 906711 · © 2013 SPIE · CCC code: 0277-786X/13/$18 · doi: 10.1117/12.2051168

Several web-based pill recognition tools have been published online in last few years. Pillbox [2], Drugs.com [3], RxList Pill Identification Tool [4], Healthline Pill Identifer [5], are some typical ones among them. Manual input is easy to use but far from fast and automatic. When a large number of pills need to be handled, this work turns out to be time consuming and costly in manpower. Manual input method is impossible to be applied to do batch processing.

Automatic method is another approach. Andreas Hartl [6] presents a mobile computer vision system. For the sake of implementation and efficiency, only pill color and contour shape have been taken into consideration. Imprint is first treated as a type of feature in the work of Young-Beom Lee [7] by means of Hu moments. However, the result of this 7-dimensional feature is far from satisfactory as shown in experiment part. Shape Distribution is introduced in [8] to measure the similarity between 3D shapes. [9] makes use of this sense and applies it to deal with pill images. Miha Mozina [10] gives another thought. Instead of using descriptors, this author utilizes a similar sense with PCA to find the first six principal components for each category, which is considered as a statistical model. Results show that it works well when detecting defects on the surface. But when being applied to do recognition, the random rotation in given pills asks the input image to be aligned into upright position beforehand.

3. IMPRINT FEATURE EXTRACTION

3.1 Imprint Extraction

In our method, imprint extraction aims to split the imprint from the raw image. Our imprint extraction is motivated by one text extraction technique: Stroke Width Transform (SWT) [11]. MSWT introduce a Switch Function to limit the search region while trying to keep as much imprint region as possible.

No matter debossed imprint or printed imprint, almost all the imprints on pills are deep-colored with respect to none-imprint region. The intuition of Switch Function is trying to split the foreground — the dark region, from the background — the bright region. Foreground function $F_i(x,y)$ and background function $B_i(x,y)$ can be seen as two functions that make up the upper and lower bound for a given image $I_i(x,y)$. Switch function $S_i(x,y)$ is generated by comparing the distance between $B_i(x,y)$, $F_i(x,y)$ and $I_i(x,y)$ as (1). The 1D case of this process can be illustrated as in Figure 2:

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

Where suffix i can be b — the switch function in bright part, or d — the switch function in dark part. δ is a small value as constraint, otherwise the Switch Function $S_i(x,y)$ may bound between 1 and 0 at regions with uniform pattern.

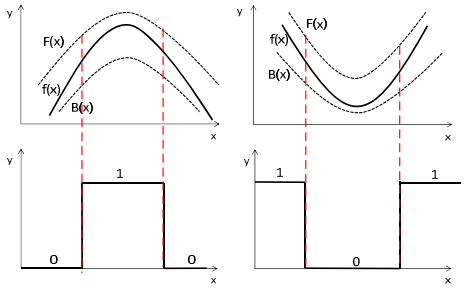

Figure 2. In 1D case, the Switch Functions on the second row are the results by comparing the $f_i(x)$ with the distance to its corresponding foreground function $F(x)$ and background function $B(x)$.

Mathematical morphology is one optional way to compute $F_i(x,y)$ and $B_i(x,y)$. Considering each stroke in imprint usually consists of both bright and dark part, we take use of Switch Function to get rough imprint regions for both two parts. For bright part, $F_b(x,y)$ stands for the morphological erosion of $I(x,y)$ and $B_b(x,y)$ for the morphological closing of $I(x,y)$. For dark part, $F_d(x,y)$ stands for the morphological dilation of $I(x,y)$ and $B_d(x,y)$ for the morphological opening of $I(x,y)$. The combination of $S_b(x,y)$ and $S_d(x,y)$ provide the restricted region for the imprint on given image.

$$S(x,y) = S_b(s,y) \oplus S_d(s,y) \quad \text{for } \forall (x,y) \in I \,. \tag{2}$$

3.2 Imprint Description

In this study, we try to treat the imprint figures as shape and the recognition then turns out to be a shape matching problem. Shape Context introduced by Belogie et al. [12] is one of the most effective shape descriptors. However, in conventional shape descriptor, shapes are generally considered as one single closed curve. While in pills, imprints are the combination of a couple of numbers, letters or symbols as in Figure 3.

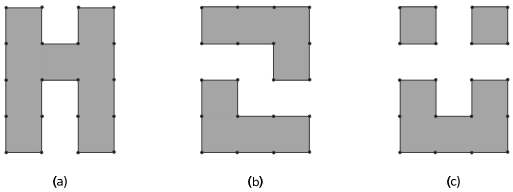

Figure 3. With the sample points on (a), (b) and (c) being the same, the binning strategy in original Shape Context is incapable of distinguishing these three shapes.

Considering this property, we suggest a weighting strategy. This weighting strategy replaces the binning method for constructing histograms in Shape Context. In binning the histogram for point p_i, instead of giving a constant score for all the points in $\{p_0, p_1, ..., p_{i-1}, p_{i+1}, ..., p_n\}$, different weights should be attached to points depending on whether they come from the same closed curve or not. The weighted shape context at point p_i is defined as a histogram h_i of the relative coordinates of the remaining points:

$$h_i(k) = \sum w(p_i, p_j), \quad \text{where } j \neq i, p_j - p_i \in \text{bin}(k) \,. \tag{3}$$

Using n_i as the number of sample points in the closed curve that p_i belongs to, n as the total number of sample points in the shape, we consider that two constraints should be attached to the weighting strategy:

1. If the n_i to n ratio close to 0, the closed curve that p_i belongs to is possibly an outlier. It should be attached with a small weight.
2. If the n_i to n ratio close to 1, it means that the closed curve p_i belongs to is a dominant component in the shape. Large weight should be given.

Having these constraints, linear weighting strategy is the most direct way to do this work, where the weight for the point pair p_i and p_j is proportional to the n_i to n ratio if they come from the same closed curve:

$$w(p_i, p_j) = \begin{cases} \dfrac{n_i}{n^2 - 2n \cdot n_i + 2n_i^2} & \text{curve}(p_i) = \text{curve}(p_j) \\ \dfrac{n - n_i}{n^2 - 2n \cdot n_i + 2n_i^2} & \text{curve}(p_i) \neq \text{curve}(p_j) \end{cases} . \tag{4}$$

$\text{curve}(p_i)$ represents the curve that p_i belongs to. $w(p_i, p_j)$ stands for the weight given to p_j when considering p_i. The denominator is a normalizing term for normalizing the built histogram h_i.

Inspired by sigmoid function, one possible strategy is that we value the pairs from the same closed curve by giving a large weight close to 1, if its n_i to n ratio not close to 0. This can be written as a transform of sigmoid function:

$$w(p_i, p_j) = \begin{cases} \dfrac{2\text{sig} - 1}{(4\text{sig} - 3) \cdot n_i + (2 - 2\text{sig}) \cdot n} & \text{curve}(p_i) = \text{curve}(p_j) \\ \dfrac{2 - 2\text{sig}}{(4\text{sig} - 3) \cdot n_i + (2 - 2\text{sig}) \cdot n} & \text{curve}(p_i) \neq \text{curve}(p_j) \end{cases}, \quad \text{where } \text{sig} = \frac{1}{1 + e^{-10 \cdot \frac{n_i}{n}}} . \tag{5}$$

The defect in sigmoid function is that for most of the ratios from 0 to 1, the weight is set close to 1. An alternative is using exponential function. It offers a smooth transition from 0 to 1 within the range of the ratio.

$$w(p_i, p_j) = \begin{cases} \dfrac{1 - \exp}{(1 - 2\exp) \cdot n_i + \exp \cdot n} & \text{curve}(p_i) = \text{curve}(p_j) \\ \dfrac{\exp}{(1 - 2\exp) \cdot n_i + \exp \cdot n} & \text{curve}(p_i) \neq \text{curve}(p_j) \end{cases}, \text{where } \exp = e^{-3.0 \cdot \frac{n_i}{n}}. \quad (6)$$

4. PILL RECOGNITION

Having a pill dataset, pill retrieval can be achieved by comparing the imprint feature of the query one with all the categories in the dataset. Then return the index of the category with smallest distance or largest similarity. This kind of recognition falls into the framework of nearest-neighbor methods, which is highly computational demanded.

For reducing the timing consumption, our consideration is making the candidate categories as fewer as possible. Hence, we try to combine color and contour shape features in the recognition system. For one input image, it will first be compared with all the categories by the feature of color and contour shape. When doing the comparison, the similarities between the query and all the categories are computed, and then sorted into order. Each comparison reserves only those with high rank in the sorted order, which means categories having large differences with the query in color or contour shape will not be considered in imprint recognition. These categories are built into a so-called short list. Then a more decent recognition algorithm can be performed on a relatively smaller dataset and speed up the whole process.

5. RESULTS

To evaluate the performance of the proposed pill recognition system, we have built an image-capturing device and constructed the captured images into two datasets.

5.1 Intensive Dataset

To evaluate the performance of the imprint feature, we manually picked out 30 different pill categories, each with its unique imprint on it. These pill categories are similar in contour shape and size. Query images are generated by randomly rotating the images, as well as randomly changing the brightness from -30% to 0%. 5 query images are generated for each category, which means there are 150 query images for testing in the dataset and at least one image as the template in each category. All the pill images are normalized into the size of 200*200. In testing intensive dataset, the pill recognition system makes use of only the contour shape feature and imprint feature, while color being neglected. Considering the similar shapes of the pills in this dataset, imprint feature plays the most significant role in the evaluation. Its performance can be well evaluated.

Figure 4. Sample pill images in the intensive dataset.

Methods	1st	2nd	3th	4th	5th
MSWT+SC	96.67	97.33	98.00	98.00	98.00
MSWT+WSC (linear)	97.33	98.67	99.33	100.00	100.00
MSWT+WSC (sigmoid)	52.00	60.67	68.00	76.00	81.33
MSWT+WSC (exponential)	98.00	98.67	98.67	99.33	99.33

Table 1. The comparison of retrieval results at top 5 ranks.

We can see that our proposed method returns a high accuracy rate. In the comparison within these weighting strategies, sigmoid weighting strategy does not perform as expected. For linear weighting and exponential weighting, each has its own merits. The failure in sigmoid may be the reason that only the points from the same closed curve can be binned into histogram, which means the whole configuration of the imprint shape lost in binning process.

5.2 Extensive Dataset

The extensive dataset is a relatively large dataset. It contains nearly 2500 different pill categories, each category providing at least one image for corresponding pill. The query images are generated in the same way as in intensive dataset. Five query images are generated by randomly translating, rotating the images. Brightness and contrast in the generated query images have also be changed from -30% to 0% randomly. The total number of query image in this test is more than 14000, 2500 for template images and more than 12000 for query images. This dataset is challenging because of not only the scale, but also a large number of pills with similar visual appearance.

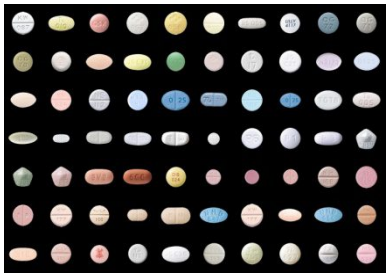

Figure 5. Sample pill images in the extensive dataset.

Method	1st	2nd	3rd	4th	5th
Shape Distribution	0.74	0.99	1.12	1.23	1.33
Hu Moment	7.85	11.91	14.55	16.46	18.21
MSWT+SC	77.87	85.08	87.91	89.52	90.86
MSWT+WSC (linear)	75.12	83.27	86.62	88.50	89.94
MSWT+WSC (sigmoid)	33.23	42.56	47.53	51.18	54.17
MSWT+WSC (exponential)	79.62	86.47	89.27	90.80	91.91

Table 2. The comparison of retrieval results at top 5 ranks.

The results show that our proposed method outperforms the existed algorithm with great improvements. In the comparison within proposed weighting strategies, exponential weighting strategy achieved the best result with 91.91% accuracy rate in the top 5 results. Linear weighting performs a bit worse than original Shape Context because it gives a small weight for most of the ratio between 0 and 0.5. The point pairs from different closed curves make more contribution and make the descriptor lose its discrimination, since our histogram is arranged in log-polar spacing.

6. CONCLUSION

In this paper, we propose one pill recognition method that outperformed those existing algorithms. The main contribution of the method is in the handling of imprint feature. In imprint extraction, Switch Function effectively eliminates the noise in result and maintains the continuity of each stroke. In imprint description, we modified the Shape Context by substituting the binning process with a weighting strategy, aiming to deal with imprint with multiple closed curves. Linear and exponential weighting strategies show their fairly good retrieval capability, while sigmoid weighting strategy performs far from our expectation. This pill recognition system can also be applied to the mass production in pharmaceutical factories as doing quality inspection. It is also believed to be an effective means for blocking the circulation of illicit drugs in our society. We are now extending the system to be able to deal with imprints with blocks and articulations, where MSWT may lose its effectiveness. What's more, we are trying to further reduce the time consumption by grouping similar categories together to construct a tree-like structure.

REFERENCES

[1] "CFR - Code of Federal Regulations Title 21," U.S. Food and Drug Adminstration, http://www.fda.gov.

[2] "Pillbox BETA," National Library of Medicine, National Institutes of Health, United States, http://pillbox.nlm.nih.gov/.

[3] "Drug.com: Pill Identifier Wizard," Drug.com, http://www.drugs.com/pill_identification.html.

[4] "RxList Pill Identification Tool," RxList, http://www.rxlist.com/pill-identification-tool/article.htm.

[5] "Healthline Pill Identifer," Healthline.com, http://www.healthline.com/pill-identifier.

[6] A. Hartl, C. Arth, "Computer-Vision based Pharmaceutical Pill Recognition on Mobile Phones" Proceedings of CESCG 2010: The 14th Central European Seminar on Computer Graphics, May 2010.

[7] Lee, Young-Beom, Unsang Park, and Anil K. Jain. "Pill-ID: Matching and retrieval of drug pill imprint images." Pattern Recognition (ICPR), 2010 20th International Conference on. IEEE, 2010.

[8] Osada, Robert, et al. "Shape distributions." ACM Transactions on Graphics (TOG) 21.4 (2002): 807-832.

[9] Caban, Jesus J., Adrian Rosebrock, and Terry S. Yoo. "Automatic identification of prescription drugs using shape distribution models." Image Processing (ICIP), 2012 19th IEEE International Conference on. IEEE, 2012.

[10] Bukovec, Marko, et al. "Automated visual inspection of imprinted pharmaceutical tablets." Measurement Science and Technology 18.9 (2007): 2921.

[11] Epshtein, Boris, Eyal Ofek, and Yonatan Wexler. "Detecting text in natural scenes with stroke width transform." Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, 2010.

[12] Belongie, Serge, Jitendra Malik, and Jan Puzicha. "Shape matching and object recognition using shape contexts." Pattern Analysis and Machine Intelligence, IEEE Transactions on 24.4 (2002): 509-522.