Technology Students' Gymkhana
Indian Institute of Technology Kharagpur

Open Soft Problem Statement

28 February, 2018

Digital Convergence : Digitization of Doctors' Hand Written Prescriptions

Origin of the Problem

Medicine becomes one of the most important parts of our life. We get this medicines through our doctors and doctors in our life is closely linked from very early of our birth to the last breath. Most of the time medicines are beneficial or at least don't harm to the patient but on occasions, they do harm the person taking them if wrong medicine or correct medicine in wrong dosage is taken. In this concern, Centre for Medicare & Medicaid Services (CMS[#1]), requested Institute of Medicine (IOM[#2]) to study the prevalence of such medication error and formulate a way to reduce it. The outcome of the report[#3] entitled "Prevention Medication Error" finds that medication errors take place surprisingly common and costly for the nation, and it outlines a comprehensive approach for decreasing the prevalence of these errors. It is also observed that, errors take place are usually during every step of the medication process namely procuring the drug, prescribing it, dispensing it, in its administration and monitoring its impact. More significantly, sloppy handwriting of a doctor found to be a major concern in this regard. In another study[#4], it is observed that sloppy handwriting causes death of approximately 7000 people yearly, it's a shocking statistic. According to July 2006 report[#5] it is observed that medication mistakes also cost more than 1.5 million person annually. Many such errors result from unclear abbreviations and dosage indication and illegible writing on some of the 3.2 billion prescriptions in US per annum.

With the above-mentioned statistics at hand, we can remark "*Thousands of people are dying although we have technology today to prevent these errors. Can this technology help us to reduce such medical errors?*"

As a step to mitigate the errors, it is proposed an approach called **DigiCon** as an initiative of digital convergence and we call it as "Digitization of Doctors' Handwritten Prescriptions".

Objectives of DigiCon

The objective in the task of DigiCon is that let allow a doctor to write his prescriptions the conventional way (i.e., using their pen and paper). From the scanned version of the prescription, a handwritten character recognition will be followed to capture the data (name of the patient, symptoms, findings, prescription of medicine, tests, advice, etc.) written by the doctor. Since, the accuracy rate of the state-of-the-art hand written character reorganization is not still up to the acceptable level, we propose to apply an error correction mechanism to reduce the errors. The solution does not oppose the age-old convention and affordable as it is mostly a software solution with a minimum hardware requirement.

Problem Statement

Input: A scan copy of a doctor's handwritten prescription and stored as an image file.
Procedure (also, see the figure in the next page):

1. *Image pre-processing:* To deal with the low quality image, noise in scan, binarization, alignment, etc. This step will produce an acceptable and processable image form.

2. *Image segmentation:* In this step, we have to segment into a number of blocks identifying different regions such as computer printed parts, sketches, computer printed images, and hand-written texts, etc. The blocks containing the hand-written texts are the regions of interest (RoIs). This step will return all the RoIs in the prescription document.

3. *Hand-written text recognition:* The RoIs involving the doctor's hand-written parts are the input in this step. For each such RoI, we have to extract words in them. Then, in each word, we have to identify the characters in it. Thus, the output in this step will be the character images in each word in each RoI.

4. *Character level recognition:* Each character image in the last step, will be processed optically to recognize the character and finally it can be stored in the form of ASCII character. The outcome of this step is the ASCII form of each word in each RoI.

5. *Sloppy hand writing correction:* The words are to be processed to check, if there is any spelling mistakes or predicting the correct words in the hand-written texts. For this purpose, you should consult a "medicine vocabulary" (such a vocabulary available in the Net and freely downloadable). Hint: You can follow any English language text prediction tool, and fine tune the same with a language resource model based on the "medicine vocabulary" as the corpus. Such a language prediction tool will predict the most probable correct words with their ranks. Output in this step will be the correct words in each RoI.

Output: The output is a document file which will place items as image in the place as it is they are there in the input prescription, except the digitized version of the doctor's hand written text with corrections.
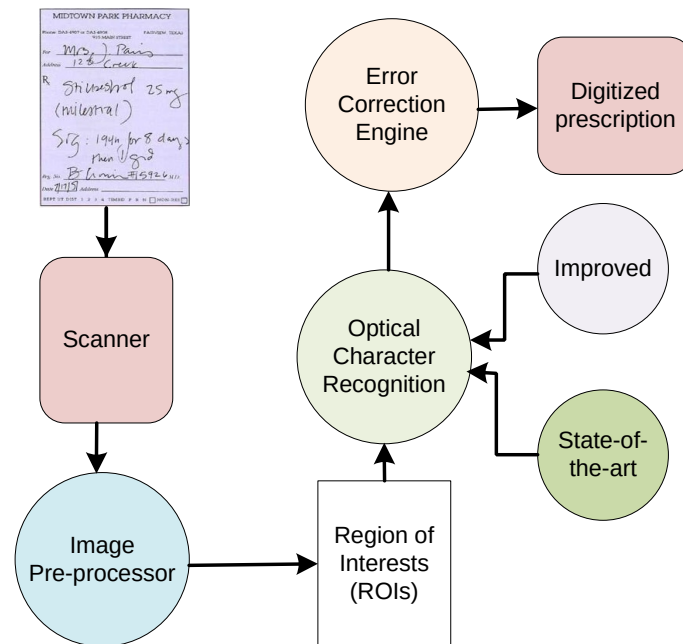
Figure: An overview of DigiCon

Submission Instruction

1. A technical report vis-à-vis a clear description of each step you have carried out. Please clearly state any reasonable assumption if you make in your implementation. The reference to all tools you have used should be stated explicitly with their sources.

2. A PDF version of the code with comments and sufficient annotation.

3. Soft copy of the source code, interface, etc. which you have developed.

4. A README file mentioning a) packages to run the software, b) makefile to build the software and hardware and OS platform to run the software.

5. A full demonstration of the software.

References:

[1] Centre for Medicare & Medicaid Services (CMS), https://www.cms.gov/
[2] Institute of Medicine (IOM), http://www.iom.edu.np/
[3] IOM Report on "Prevention Medication Error, July 2016"
http://www.nationalacademies.org/hmd/~/media/Files/Report%20Files/2006/Preventing-Medication-Errors-Quality-Chasm-Series/medicationerrorsnew.ashx
[4] http://content.time.com/time/health/article/0,8599,1578074,00.html
[5] https://www.fda.gov/drugs/resourcesforyou/consumers/ucm143553.htm
[6] OCR : Find many tool freely available in the Internet as open source
[7] CMU tool kit : It is freely available

[8] Language corpus : Medicine vocabulary of different sizes are available in the Net.
[9] Language prediction tool. Possible available in the Internet as well.