Positive identification of Chinese characters from images

Author: Tommaso Torelli, tomukun@stanford.edu

Android camera phone utilization: optional, under a client-server type setup

Desired features

1) Ability to extract Chinese characters even in the presence of non-uniform backgrounds (restaurant menus, shop signs, etc.)

2) Ability to handle characters of different colors within the same image (as long as they are sufficiently distinct from their immediate background)

3) Ability to handle different font types

4) Ability to look up character meaning not just of single characters, but small character groupings against a variety of known character databases

5) Does not use conventional OCR approaches

Assumptions

1) Character sequences must follow either one of these two linear arrangements (do not support a mixture of both for the same image → this is rarely observed):
   - Horizontal arrangement: from left to right first and then top to bottom
   - Vertical arrangement: from top to bottom first and then left to right

2) A fixed rectangular space, which I will refer to as a block, is allocated for each individual character, regardless of its complexity or stroke density

3) Contiguous lines of characters must be separated by uniform spacing where no individual characters are allowed, although no assumptions are made as to the background uniformity of such spacing. The same consideration applies to the spacing between contiguous character blocks within a line.

4) The input images must be from a view orthogonal to the plane of the text, although a deviation of up to ±15° from the orthogonal can be tolerated. The inevitable loss of character detail resulting from a skewed perspective is something that can in no way be obviated, so this is limitation is fairly benign. No assumptions are needed as to the orientation of the text in the plane, other than it should be arranged in a Cartesian fashion.
5) All strokes within each character must have the same color

6) Character composition must follow one of the schemes identified in reference [1]

7) The text must have sufficient character density, i.e. performance is expected to be poor on

**Description**

The processing will be subdivided into 4 distinct phases. Phase 1 is performed on the entire image, whereas successive phases are performed only on individual character blocks, i.e. small rectangular portions of the entire image.

❖ Phase 1 – Character layout of text in the image

   1) Perform DCT analysis to determine:
      a. Off-axis text orientation (from horizontal/vertical)
      b. Character line arrangement after rotation (whether horizontal or vertical)
      c. Character block size in x and y
   2) Use image projection data along each character line to identify and extract individual character blocks of the predicted size. After this step, the rest of the processing is done on individual character blocks.
   3) Foreground polarity detection: compare the brightness of spacing between character lines and blocks, vs. average brightness computed over a handful of blocks to determine whether

❖ Phase 2 – Image pre-processing (individual blocks)

   If block size determination can be made to work accurately, only a modest amount of pre-processing should be required. The operations are akin to those used in the Spring 2012 Mobile Chinese Translator project [2], with some notable differences:

   1) A median filter with small kernel size (3x3 or 5x5 at most) to smooth out very high-frequency (salt-and-pepper) noise that may be present in the image. This should still preserve the characters’ edges, as long as they span multiple pixels
   2) Unsharp filter for image de-blurring should also be useful

   Conversion to gray scale will not be performed, as color information can be exploited to improve subsequent image thresholding; also, contrast enhancement should not be needed.

❖ Phase 3 – Block binarization by Otsu’s method

   The power of extracting individual character block lies in the fact that even a simple approach such as Otsu’s method should adequately handle character binarization. One possible refinement that may be desirable for highly colorful images would be to convert each block
image to different color spaces (YUV, HSV, etc.) and look at binarization for all of the different components and merge the binary masks possibly via a majority voting scheme.

- Phase 4 – Connected component analysis of the binary image

Identify all of the connected components in each block and attempt to match each of them to a Chinese radical as listed in Ref [1] based on extracted region attributes. The match should accommodate for various font variations and carry a quality score associated with it.

If 2 or 3 radicals that compose each character can be identified with sufficiently high confidence, the character meaning can be looked up against a known list of characters organized by radical composition. This is by far the most efficient way of achieving semantic interpretation of the characters.

Even for characters where maybe only 1 or no radical can be identified, a second pass analysis of character groups (pairs, triplets, etc.) could be attempted and compared against a table of such high frequency groupings. The “gap” character(s) in the grouping could then be positively ID’ed by performing an image difference against potential multiple candidates and picking the candidate with the smallest difference.

References

[1]: Chinese characters decomposition WikiCommons Project (http://commons.wikimedia.org/wiki/Commons%3aChinese_characters_decomposition)
