International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 3 | Issue 5 | ISSN : 2456-3307

Detection and Recognition for Reading Text in Images

Pooja Kumari[1], Mrs.Mamta Yadav[2]

[1]M.Tech Scholar CSE, M.D.U Rohtak, YCET Narnaul, Mahendergarh, India
[2]Assistant Professor CSE, M.D.U Rohtak, YCET Narnaul, Mahendergarh, India

ABSTRACT

Detection And Recognition for Reading Text in Images is a difficult but important problem. It can be summarized as: how to enable a computer to recognize letters and digits from a predefined alphabet, possibly using contextual information. Various attempts at solving this problem, using different selections of features and classifiers, have been made. Human performance has been achieved in accuracy by automated text recognition systems, and has been bypassed in speed for the case of single size, single font, high quality, known layout, known background, text. When one or more of the above parameters are changed, the problem becomes increasingly difficult. In particular, attaining human performance in recognizing cursive script of varying size, varying style, unknown layout, unknown background is far from the reach of todays' algorithms, despite the continuous research effort for almost four decades. In this report, we analyze the problem in detail, present the associated difficulties, and propose a coherent framework for addressing automated text recognition. A lot of people like to say that the world is overwhelmed with information that is still harder and harder to deal with, both for individual humans living in the overwhelmed world and for the technology they use. Popularity of mobile devices equipped with cameras has influenced peoples' lives in many ways recently. One of these changes is that people started to take photos as notes about things which are not of visual nature as opening hours or traffic schedules. Taking a picture of the signs became a very convenient way of storing such information, however later retrieval of such "photographic notes" with any meta-data may became very time consuming.

Keywords : Detection and Recognition, ASCII Code, Optical character Recognition, Hausdorff Distance, Euclidean Distance

I. INTRODUCTION

The field of automated pattern recognition was originally stimulated by studies in optical character recognition. Optical character recognition means the ability of mappin grey-level images of characters into equivalent ASCII code'. The achievement of this goal implies the automation of time-consuming - yet important - tasks such as data entry, check processing and mail interpretation. However, the simplicity of the problem statement belies the complexity of the problem. This complexity is due - mainly - to two factors: the variability in writing styles and sizes across different writings, and the a priori uncertainty in the text layout. Two major approaches to the problem can be identified in the literature: the statistical approach and the linguistic approach. However, neither approach has yielded the desired performance needed in the above cited applications, especially if the documents are handwritten. The main difficulty encountered in both approaches is the question of representation, or feature selection. What is a character? a binary image? a collection of strokes? or a set of numbers resulting from morphological operations? Apparently, this question can only be answered through a consideration of the text

generation process and the structure of the uncertainty present.We propose applying an information theory framework to understand the problem and try evaluating previous approaches within such a framework. The second major encountered difficulty is the question of similarity. What similarity measure best models our perception? A weighted Euclidean distance? A Hausdorff distance? or a set of rules? It is true that the questions of representation and similarity are closely tied. Nevertheless, we argue that the classical notion of looking for a metric (in the mathematical sense) is not the proper notion for recognition.

II. DETECTION AND RECOGNITION FOR READING TEXT IN IMAGES ALOGRITHM AND THEORM

In this context, y represents the grid of squares, and x is of course the observed image. The usual discriminative Markov field model (repeated here for convenience)

p (y | x, θ, I) ≡ 1Z (x) exp (X C∈C UC (yC, x; θC)

will employ two types of compatibilities. One is local, relating image features to a label for a particular region, and the other is contextual, relating image features to the labels for pair of neighboring regions. Both of these take the usual linear form

Ui (yi , x, θi) = θi (yi) · Fi (x), i ∈ V (3.2)

Uij (yi , yj , x, θij) = θij (yi , yj) · Fij (x), i ∼ j, (3.3)

where i ∼ j indicates that regions i and j are neighbors in the grid. we will use the laws of probability to incorporate all of the spatial information, while leaving the mixed regions unlabeled. This allows the model to form its own beliefs about whether such regions should be labeled Sign or Background, without attempting to train it to take an unnaturally rigid stance on such data. A na¨ıve approach for giving the mixed regions no labels is to literally omit them from the training data. This is reasonable for a local model, where predictions are made independently, but it is the wrong approach once spatial dependencies are introduced. In the grid for the image, the mixed regions would simply be removed from the graph, eliminating all compatibility functions involving such regions. This will be problematic because features and nodes at the interface of Background and Sign regions are discarded and the model cannot learn the properties of the transitions between different region types.

Properly training the model to incorporate the spatial dependencies with incomplete labels involves using a marginal conditional likelihood in the parameter posterior. If Dy only contains some of the labels for the images Dx, then there are “missing,” unobserved labels Du that must be accounted for. These are the yellow regions Recall that maximizing the parameter posterior p (θ | D, I) involves the likelihood

p (Dy | Dx, θ, I). By the product rule of probability, we have that

p (Dy | Dx, θ, I) = p (Dy, Du | Dx, θ, I)

p (Du | Dy, Dx, θ, I)

Domain Deformation : Theory

Assume that under hypothesis Hi:

g (t) - hi (x (t)) - (hi 0 X) M

where x(t) is an order-preserving homeomorphism of the unit interval I onto itself. We will show later that a solution x for the above equation exists if and only if g and hi have the same sequence of extrema. The reader should note that, throughout the coming discussion, we will be dealing with three different spaces: X, H, and W. First, we will define X, and then we will present a few lemmas to gain some understanding about the space X.Let X be the space of all order-preserving homeomorphisms of the unit interval I onto itself. Then, the following two lemmas are well known.

Lemma 3.1 A Junction x is an element of X if and only if it can be represented as a continuous, strictly

increasing, Junction joining the origin to the point (1,1) of I x I (see Figure 3-2). The horizontal axis in Figure 3-2 (also called the t-axis) Z's the domain of g, and the vertical axis (also called the x-axis) Z's the domain of hi.

Lemma 3.2 The pair (X, 0) ,is a group.Note that the inverse x-I is a reflection of x around the diagonal of I x 1.

Lemma 3.3 The space X, viewed as a set, is convex. Proof: By convex we mean, if xi and X2 belong to X, then so do ax, + (1 - a)X2 for all a in [0, 1]. Let x = ax, + (I - a)X2. Then x is continuous, being a weighted sum of continuous functions. Letting t, < t2, we get
X(t2) = aXl(t2) + (I - a)X2(t2)
> OZXI(tl) + (1 - Ce)X2(tl) X(ti)

Finally, x(O) - 0, and x(l) - 1. Hence, x is in X.
Next, we define the second space of interest, H, and its subspace, Hg:
Definition 3.4 Let H be the space of real functions of bounded variations defined over 1. Let g be an element of H. Hg is defined to be the set of all functions in H which can be obtained from g through order-preserving, homeomorphic, domain deformation. In other words,
Hg - f f : g - f 0 X, X G XI

Note that we already have an onto mapping from X to Hg, mapping x to g o x-'. However, this mapping is not one-to-one in general. Consider, for instance, the constant function g(t) -_ 1. Then, Hg is the singleton f gj, since I o x-' - I for all x in X. By removing from X the "redundant" x's, the onto mapping becomes a bisection. For the case g = 1, the redundant deformations are all the deformations except x -_ i. The following lemmas characterize the redundant domain deformations for a general function g.

Lemma 3.5 Let f and g be both strictly increasing (decreasing) functions. Assume that the relation 9 = f o x has a solution in X. Then, that solution is unique. Proof: Assume the contrary. Let xi and X2 be two different solutions of the equation g - f ox. Then, there exists a point a at which they differ. Let x (a) - bi, X2 (a) = b2 - However, g(a) -_ f [xi (a)] = f [X2(a)]. Therefore, g(bl) =: g(b2) for b, : b2. This is a contradiction, since g is strictly monotonic. Hence, the solution x is unique. Lemma 3.6 Let f , g G H . Let 9 - f ox has a solution in X. Then, x Z's a bisection between the local maxima(minima) of g and the local maxima(minima) of f.
Proof: Let a be a point at which g has a local maximum. Then, there is an open nbhd of a, B(a) C 1, such that g(t) < g(a) for all t in B(a)' Consider x[B(a)]. It isopen since x is a homeomorphism, and contains x(a). For all u in x[B(a)], we have f (u) < f [x(a)], otherwise x-'(u) would be a point in B(a) such that 9[x-'(u)] > g(a). Hence, x (a) is a local maximum of f . Similarly, if b is a local maximum of f , we can prove that x-'(b) is a local maximum of g. Finally, the minima can be treated in an analogous way.

Character Overlap

A pair of neighboring spans may either overlap or have a gap between them. In the case of overlap, a simple energy term is added:
U O n,r θ O= θ O n,r, (4.3)
depending on how many pixels overlap—from n to r—between the spans. Using this information, we allow character bounding boxes to overlap (as in the example filigature of Figure 4.3), but the degree of overlap allowed is soft and flexible

Character Gap

As stated above, a pair of neighboring spans may also have a gap between them.For instance, in Figure 4.3, the character i is separated from g by a few pixels.) In this case, the gap is scored by a learned compatibility function
U G n,r x; θ G= Xr i=n θ G · Fi (x), (4.4)

Lexicon Information

Our model also features a parameter that will allow a bias for character sequences that compose a lexicon word,

U W θ W = θ W . (4.5)

When calculating the total score (as detailed in the next section), this term will only be included in portions that are part of a lexicon word normalizing the feature vector element by its

Model Inference

The recognition task can be thought of as finding the segmentation and corresponding labeling that maximizes a total (summed) score. The constituents of this score—the exponent of the exponential model—were described in the previous section. Here we describe how to find the segmentation and labeling that gives the best overall score. The inference process can be accomplished in a model with or without the use of a lexicon. Omitting lexicon information is simpler, so we begin by describing a model without

4.2 Markov Models for Recognition

Just like the discriminative Markov field for detection in Chapter 3, a similar model for recognition involves defining parameterized compatibility functions for the data and the labels. For the recognition problem, the model input will be size-normalized character images and the output is the predicted character labels. In this section we will outline the details of our model, including the form of the input and features, the relevant information being utilized, and the particular compatibility functions that are learned to form the model.

Our model and the subsequent experiments make the following assumptions:

1. The input is all of the same font
2. Characters have been segmented (that is, the coordinates of their bounding boxes are known), but not binarized
3. Word boundaries are known

III. Uses

The advantages of OCR are numerous, but namely it increases the efficiency and effectiveness of office work. The ability to instantly search through content is immensely useful, especially in an office setting that has to deal with high volume scanning or high document inflow. You can now use the copy and paste tools on the document as well, instead of rewriting everything to correct it. OCR is quick and accurate, ensuring the document's content remains intact while saving time as well. When combined with other technologies such as scanning and file compression, the advantages of OCR truly shine. Workflow is increased since employees no longer have to waste time on manual labor and can work quicker and more efficiently.

IV. REFERENCES

[1]. Agarwal, Shivani, Awan, Aatif, and Roth, Dan. Learning to detect objects in images via a sparse, part-based representation. IEEE Transactions on Pattern Analysis and Machine Intelligence 26, 11 (2004), 1475–1490.

[2]. Baird, Henry S., and Nagy, George. Self-correcting 100-font classifier. In Proc. of SPIE: Document Recognition (1994), Luc M. Vincent and Theo Pavlidis, Eds., vol. 2181, pp. 106–115.

[3]. Bapst, Fr'ed'eric, and Ingold, Rolf. Using typography in document image analysis. In Electronic Publishing, Artistic Imaging, and Digital Typography (1998), vol. 1375 of Lecture Notes in Computer Science, pp. 240–251.

[4]. Bargeron, David, Viola, Paul, and Simard, Patrice. Boosting-based transductive learning for text

detection. In Proc. Intl. Conf. on Document Analysis and Recognition (2005), pp. 1166–1171.

[5]. Bazzi, Issam, Schwartz, Richard, and Makhoul, John. An omnifont openvocabulary OCR system for English and Arabic. IEEE Transactions on Pattern Analysis and Machine Intelligence 21, 6 (1999), 495–504.

[6]. Beal, Matthew J. Variational Algorithms for Approximate Bayesian Inference. PhD thesis, University College London, London, 2003.

[7]. Beaufort, R., and Mancas-Thillou, C. A weighted finite-state framework for correcting errors in natural scene OCR. Proc. Intl. Conf. on Document Analysis and Recognition 2 (2007), 889–893.

[8]. Belongie, S., Malik, J., and Puzicha, J. Shape matching and object recognition using shape contexts. IEEE Transactions on Pattern Analysis and Machine Intelligence 24, 4 (2002), 509–522.

[9]. Berger, Adam L., Della Pietra, Stephen A., and Della Pietra, Vincent J. A maximum entropy approach to natural language processing. Computational Linguistics 22, 1 (1996), 39–71.

[10]. Bernstein, Elliot Joel, and Amit, Yali. Part-based statistical models for object classification and detection. In Proc. Conf. on Computer Vision and Pattern Recognition (2005), pp. 734–740.

[11]. Bledsoe, W. W., and Browning, I. Pattern recognition and reading by machine. In Proc. of Eastern Joint Computer Conf. (1959), pp. 225–232. 134

[12]. Blum, Avrim, and Langley, Pat. Selection of relevant features and examples in machine learning. Artificial Intelligence 97, 1-2 (1997), 245–271.

[13]. Boyd, Stephen, and Vandenberghe, Lieven. Convex Optimization. Cambridge University Press, 2004.

[14]. Brakensiek, Anja, Willett, Daniel, and Rigoll, Gerhard. Improved degraded document recognition with hybrid modeling techniques and character n-grams. In Proc. Intl. Conf. on Pattern Recognition (2000), vol. 4, pp. 438–441.

[15]. Breuel, Thomas M. Classification by probabilistic clustering. In Proc. Intl. Conf. on Acoustics, Speech, and Signal Processing (2001), vol. 2, pp. 1333–1336.

[16]. Breuel, Thomas M. Character recognition by adaptive statistical similarity. In Proc. Intl. Conf. on Document Analysis and Recognition (2003), vol. 1, pp. 158– 162.

[17]. Buntine, W., and Weigend, A. Bayesian back-propagation. Complex Systems 5 (1991), 603–643.

[18]. Carbonetto, P., de Freitas, N., and Barnard, K. A statistical model for general contextual object recognition. In Proc. European Conf. on Computer Vision (2004), vol. 1, pp. 350–362.

[19]. Caruana, Rich. Multitask learning. Machine Learning 28, 1 (1997), 41–75.

[20]. Chen, Datong, Odobez, Jean-Marc, and Bourlard, H. Text detection and recognition

[21]. in images and video frames. Pattern Recognition 37, 3 (2004), 595–608.

[22]. Chen, Xiangrong, and Yuille, Alan L. Detecting and reading text in natural scenes. In Proc. Conf. on Computer Vision and Pattern Recognition (2004), pp. 366–373.