9
Conclusion and Future Scope

"No research is ever quite complete. It is the glory of a good bit of work that it opens the way for something still better, and this repeatedly leads to its own eclipse."

Mervin Gordon

The objective of the work done in this thesis was to develop methods that can help automate processing for the modern (Intelligent Character Recognition) ICR systems. We developed new methods for two major tasks: (i) Segmenting handwritten annotations from printed documents, and (ii) Identifying the writer of handwritten words. In this chapter, the main findings with regard to the research objectives and the strengths of the proposed methods are summarized in Section 9.1. Furthermore, Section 9.2 presents a summary of the experimental results and Section 9.3 highlights the limitations and suggestions for future research.

9.1 SUMMARY OF THE PROPOSED METHODS

Due to incongruity in the properties of handwritten and printed text, OCRs use separate processing steps for the two types of text. This indicates the requirement for the separation of handwriting from the printed text. The problem of segmentation of annotations from the printed text is solved under restricted conditions. Applications dealing with retrieval and search have been effective on document images with constraints on the machine printed text.

Handwriting styles vary greatly from person to person. The preferred ways and styles of annotations also vary from person to person. Even a single document can have varied annotations. For example, they can include marks, cuts, underlined text, characters, single and multiple words, overlay text and special symbols, irregularly written text, along with the regular handwritten text. Apart from varied annotations, there exists a large diversity in document layouts. For commercial work-flow environments, the recognition systems have to be very fast. The system should be trainable to handle the large variety and it is desirable that the amount of training data required should be less so as to enable efficient automation. According to the literature, much attention has been given to develop methods that can extract text annotations. Documents that are handled are structured or semi-structured, having mostly homogeneous layouts. Such an environment is termed as a controlled environment where we restrict the location of the annotation. In contrast, there are also documents which are structure-free and have non-predictable layouts. Annotations on these documents are marked in an unstructured way which results in an This is the unconstrained, non-controlled environment. Examples of unconstrained annotations are writing within the margins, between the paragraphs, multi-oriented text-lines, overlapping with the printed text, and presence of symbolic annotations like arrows, underlines, cuts, and encirclement. Consequently, extracting multi-oriented handwritten annotations in non-predictable layouts

remains a difficult task. There is a need for robust multi-way extraction of handwritten content from heterogeneous layouts.

Sometimes, it may be of interest to know who signed or edited a document or a specific keyword by a particular author. Handwriting as a personal biometric is considered to be unique to a person. A writers individuality rests on the hypothesis that every individual has a consistent handwriting which is distinct from the handwriting of another individual. However, writer identification is a difficult problem because of the variability in the handwriting of a given writer, limited availability of labelled training data, presence of noise, etc.

The major objectives of the thesis are:

1. Extraction of multi-oriented annotations in uncontrolled environments.
2. Extraction of a specific type of annotation in a multi-oriented annotated environment.
3. Writer identification for handwritten words.

The research work for achieving these objectives was organized into the following work elements.

I. Comprehensive study of the state-of-the-art methods for printed and handwritten text, and off-line writer identification.

   We highlighted the diversity of the domain of annotation separation from the printed text. We enlisted the distinctiveness in the shape and statistics of the printed and handwritten content. We reviewed the state-of-the-art methods for handwritten and printed text separation, covering aspects such as their characteristics, constraints, and datasets used. The review is organized along six levels of segmentation involving pixels, characters, blocks, connected component, text-line, and words. It also incorporates a review of various existing post-processing strategies that can enhance the performance of the handwriting segmentation framework.

   In a similar manner, the offline writer identification is scrutinized and elaborated. This review is broadly organized along recognition granularity, such as character/grapheme, document, word, connected component, and text-line level.

   The review of past work related annotation extraction and writer identification has been presented in chapters 2 and 3 respectively.

II. Extraction of multi-oriented annotations in non-controlled environment.

   In Chapter 4, we proposed a method for multi-oriented handwritten annotation extraction using spectral partitioning. We have applied our approach on annotations written on documents such as conference papers, articles, books, office documents etc. Such documents can have non-predictable layouts and the annotations are multi-oriented and irregular, and include marks, cuts, underlined text, characters, single and multiple words, overlay text and special symbols along with the regular text. Consequently, extracting multi-oriented handwritten annotations in a real environment remains a difficult task.

   We also developed a novel discriminating feature called Envelope Straightness to separate printed text and complex annotations. Since complex annotated datasets were not available, hence we created our own dataset of 40 document images and added some complex cases of annotations like cuts, crosses, underlines, overlay text, special symbols, digits along with

the regular handwritten text. We also performed similar experiments on IAM dataset and compared our results with the use of alternative feature sets proposed in the literature.

III. Extraction of a specific type of annotation in non-controlled environment.

To extract specific annotations, we treat different types of annotations as different objects and deploy top-down visual saliency methods for fast localization of annotated regions in an image. Our first approach was supervised approach that trained a CRF to infer the saliency value for the image patches. Our second approach was a weakly supervised approach based on a formulation of *discriminant saliency*. The formulation was motivated by the Barlow's principle of suspicious coincidences, which means that the features that are present in the target class and absent when the target is not present, are most important from the saliency point of view. In other words, salient features are the ones for which the presence of the feature coincides with the presence of the object, i.e. creates suspicion about the presence of the object.

In Chapter 5, we described the first approach. The CRF model made use of a sparse encoded representation of image patches. For training, we adopted a method that could jointly learn the dictionary and the CRF parameters. In Chapter 6, we described the second approach. We reviewed discriminant saliency, its computational models and its suitability for fast localization of specific types of handwritten annotations.

Due to non-availability of the datasets, we created one in our laboratory. We generated a dataset of 360 document images annotated in free-hand with different inks by 60 research scholars. The ground truth is acquired at the bounding box level and also at the pixel level. To validate the robustness and generality of our methods we applied our trained saliency models to locate the annotations on 100 images of IAM dataset and 100 images of the PRIma-NHM dataset.

IV. Writer identification of the handwritten words

Much of the existing work on writer identification has focused on using a controlled vocabulary, i.e. the system is trained and tested on specific words. In this thesis, we developed a method in which the handwritten words used for testing need not be the same as the handwritten words used for training. Our system is applicable to documents where multiple authors have annotated the same page. The extracted features were clustered and writer-specific classifiers were then trained on each cluster. We made use of a sliding window to extract features at grapheme level. The features were clustered using k-means clustering. We envisaged that allographs that are similar would get grouped into the same cluster. The set of allographs that belong to a cluster may be coming from samples by several writers. For each cluster, we trained a suite of one-vs-rest SVM classifiers using samples that belonged to that cluster. The feature vector extracted for each word segment (window) is associated with the closest cluster and is classified by the suite of SVMs trained for that cluster. A majority voting among the classification decisions for all the windows gives the final writer label assigned to the word. In our work we adopt a 17 dimension feature set as Pixel Density, center deviation from lower baseline, Zernike moments, mean and standard deviation of vertical and horizontal projection normalized by the image width, mean of derivate of vertical and horizontal projection vector profile of the image, mean and standard deviation of vertical and horizontal runs of the image, and topological masks matching counts. Classification was done by a suite of one vs rest SVMs. In the end, majority voting was used to decide the author of the given handwritten word. We also compared feature clustering applied to graphemes with feature clustering applied to segmented characters in terms of performance for word-level writer identification.

9.2 SUMMARY OF THE EXPERIMENTAL RESULTS

One purpose of this study was to assess a new way to identify the writer of the given text word. This includes finding of a novel method to separate the core-zone from the ascenders and descenders in a word.

The main motive to explore segmentation of the respective core zone is to locate the lower baseline in a word. The center deviation of the center of mass from the extracted baseline can be used as a feature for writer identification process. With the following intention, we acclaimed zone extraction as a significant preprocessing step in handwriting analysis.

Our work in chapter 7 presents this new method for separating ascenders and descenders from an unconstrained handwritten word and identifying its core-region. The method estimates correct core-region for complexities like long horizontal strokes, skewed words, first letter capital, hill and dale writing, jumping baselines and words with long descender curves, cursive handwriting, calligraphic words, title case words, and very short words. It extracts two envelops from the word image and then selects sample points that constitute the core region envelop. The method is tested on CVL, ICDAR-2013, ICFHR-2012, and IAM benchmark datasets of handwritten words written by multiple writers. We also created our own dataset of 100 words authored by 2 writers comprising all the above-mentioned handwriting complexities. Due to non-availability of the Ground Truth for core-region extraction, we created it manually for all the datasets. In totality, we experimented on 17100 words written by 802 writers and extensively compared our work with the *state-of-the-art*.

1. For our graph-cut based framework for annotation extraction we performed two sets of experiments with different datasets and compared our work with *state-of-the-art* [Peng *et al.*, 2013; Benjlaiel *et al.*, 2014]. The first set of experiments were performed on 40 document images of our dataset. For annotations (handwritten text), the features used in this work have given a better precision of 85.40% and recall of 29.70% compared to a precision of 80.87% by [Peng *et al.*, 2013] and 53.40% by [Benjlaiel *et al.*, 2014]. For printed text, our features have a higher recall of 98.39% and comparable precision of 81.79% when compared to the features used in [Peng *et al.*, 2013] and [Benjlaiel *et al.*, 2014].

   The second set of experiments were performed on 40 document images of IAM dataset. A similar performance was shown by our method on the IAM dataset with a recall of 81.89% and precision of 97.95% on printed text and a recall of 95.87% and precision of 69.67% on handwritten text. Moreover, the inclusion of a new feature *envelop straightness* enhances the discriminability of the proposed method by 2% to 3% for both the datasets. In addition to this, we also examined general document clustering methods such as, hierarchical clustering and partitional clustering. On experimentation, it is observed that spectral partitioning performs much better than the other clustering techniques. It is because annotations on a document usually form a non-convex set of features and hence we require a way that can easily separate such intertwined spirals in the feature space.

2. The extracted handwritten regions were then analyzed for extracting the core-region of the words and identifying the writer. Good results were obtained by our method of core-region extraction to separate out the ascender and descender zones. The method was tested on CVL, ICDAR-2013, ICFHR-2012, and IAM benchmark datasets of handwritten words written by multiple writers. Our work reported an accuracy of 90.16% for correctly identifying all the three zones on 17,100 Latin words written by 802 individuals. It is worth mentioning that the reference lines obtained by other techniques are straight lines, where as, our method specifies the core region as bounded by upper and lower envelops and therefore yielded higher accuracy. Comparison was done with the *state-of-the-art* methods [Bozinovic and

Srihari, 1989; Vinciarelli and Luettin, 2001; Blumenstein *et al.*, 2002; Rehman *et al.*, 2009; Papandreou and Gatos, 2014]. An improved performance in the range of 5% to 30% was observed over the other methods.

3. In order to assess the effectiveness of the proposed approach for writer identification, we performed a series of experiments on the CVL dataset. The assessment was done a documents by 10 writers from CVL dataset. The dataset included handwritten words from a set of 4 documents from each writer. We used words from 2 of these documents for training and words from the remaining two for testing. The grapheme level clusters are trained with 140 different words and further tested on 150 different words for each writer. We conduct two sets of experiments by building the feature clusters with overlapping and non-overlapping windows. The character level feature clusters are learned from 5698 characters from 10 writers and were tested on 150 different words for each writer. Working at allograph level we assessed the writer identification rate for graphemes based clusters. We observed an of identification rate from 24.47% to 66.40% for a window size of 20 which was experimentally determined. In a similar manner, we assessed the writer identification rate for character-based clusters. We observed an identification rate from 14.09% to 18.81% for a window size of 20. The best results were obtained for overlapping windows. Possibly the non-overlapping windows result in clusters that are compact and hence are unable to separate the writers properly in the feature space. The classification accuracy is observed to be somewhat invariant to the number of clusters used for both the grapheme and character level allographic features. In the overlapping setting, more windows are generated which increases the allographs extracted from the sample words. It is also found that the grapheme level outperforms the character level analysis irrespective of window size. Our methodology (using graphemes) which is a discriminative approach presented an improvement in identification accuracy over the generative approach described in [Slimane and Margner, 2014] that uses window-based features and models a GMM for each writer. This improvement in the identification accuracy ranged from 19% to 63% at grapheme level for overlapping windows and ranged from from 18% to 22% for non-overlapping windows.

4. For visual saliency framework to extract specific annotations we performed four sets of experiments on our dataset and compared our work with a discriminating SVM classifier. In the first set, we localized *all* types of annotations on the model that is trained on images comprising all annotations. The CRF saliency model produces a recall of 0.71 for annotations and a precision of 0.92 for printed text, while the DS model produces 0.58 recall and 0.82 precision for annotations and printed text respectively. This signifies the fact that there is a significant improvement of 13.2% in the recall and 10.11% in precision when supervised saliency is used for multi-oriented annotation extraction. In the second set, we localized *specific* annotations in a *multi-annotated* document. In this setting, the model is trained with the images comprising only individual annotations. We present results on four types of annotations: underline, Marginal Text, Encircled and Inline. For underlined annotations, CRF model achieves recall of 0.75 and a precision of 0.51, while the DS model achieves recall of 0.81 and precision of 0.47. For marginal text annotations, the CRF model computes recall and precision values as 0.52 and 0.83, while the DS model produces 0.79 and 0.70 as recall and precision rates. In a similar manner, for encircled annotations, the CRF model produces recall and precision of 0.64 and 0.45 while the DS model produces the recall and precision as 0.84 and 0.63 respectively. For inline annotations, the CRF model achieves a recall of 0.51 and precision of 0.84 while the DS model computes recall as 0.18 and precision as 0.16 for inline annotations. The following results signifies that weakly supervised methods shows better results than the supervised method to recall annotations. Evidently, discriminant saliency shows an increase of 5% to 27% in the recall rates for annotations like marginal text, underlines and encirclements. However, to locate the inline annotations the CRF model in

comparison to discriminant saliency presents impressive results. It produces an increment in the recall by 26% and 30% increment in precision rates. In the third set, we localized *textual* and *symbolic* annotations separately in a multi-annotated document. In this setting, the model is trained with the images comprising only individual annotations. For such set of experiments discriminant saliency shows impressive results in comparison to CRF and produces a recall and precision of 0.57 and 0.78 to locate textual annotations. It also produces a recall of 0.71 and precision 0.83 of to locate symbolic annotations on a multi-annotated document. In contrast to supervised CRF saliency model the recall and precision rates for textual annotations are 0.57 and 0.89 and for symbolic annotations the recall and precision are 0.21 and 0.83 respectively. Finally, in the fourth set, we localized *specific* annotations in a *single-annotated* document. For underlined annotations, CRF model achieves recall of 0.81 and a precision of 0.88, while the DS model achieves a recall of 0.68 and a precision of 0.94. For marginal text annotations, the CRF model computes recall and precision values as 0.46 and 0.78, while the DS model produces 0.91 and 0.87 as recall and precision rates. In a similar manner, for encircled annotations, the CRF model produces recall and precision of 0.64 and 0.93 while the DS model produces the recall and precision as 0.87 and 0.70 respectively. For inline annotations, the CRF model achieves a recall of 0.75 and precision of 0.83 while the DS model computes recall as 0.49 and precision as 0.52 for inline annotations. The results depicts that for encirclement and marginal annotations discriminant saliency produces better performance in contrast to CRF model. It produces an increment of 22% in recall for encirclement annotation and 45% increment in recall for marginal annotations. Moreover, for underlined and inline annotations the CRF model gives an increase of 12% and 26% in recall rate against discriminant saliency. We also presented the effectiveness of our work on the benchmark datasets. We located textual annotations on 100 IAM and PRImA NHM Dataset images. We obtained a recall of 0.68 and precision of 0.99 on IAM dataset and a recall of 0.87 and precision of 0.86 on PRImA NHM dataset using CRF supervised saliency model. A similar performance was shown by discriminant saliency with a recall of 0.98 and precision of 0.99 on the IAM dataset and a recall of 0.77 and precision of 0.66 on PRImA NHM dataset. In comparison to SVM for *specific* annotation localization, our saliency methods gives better performance. There is nearly an increment of 20% to 40% in overall recall and precision rates for both CRF and discriminant saliency models.

9.3 FUTURE SCOPE

The main objective of the research work undertaken in this thesis is to enhance the performance of an annotation extraction system for a multi-oriented and multi-layout environment and to identify the writer for a specific handwritten word. Numerous research facets emerging out of this work may provide worthwhile exploration directions to researchers for their endeavors. Some of the main directions can be given as follows.

I. Despite decades of research in annotation extraction, it is found that most of the literature concerns with the problem of text extraction and classifying as printed or handwritten. However, annotations broadly include anything drawn or written by hand. There can be cuts, crosses, arrows, underlines, pictures, flowcharts, inline text, and special symbols. Although few methods like [Peng et al., 2013; Seuret et al., 2014] are there that have considered the overlay text and have solved the concerned problem at the pixel level, nevertheless their datasets were deprived of other annotations. Annotation extraction at character, word, line, or connected component is only limited to text. One of the tough challenges for all the researchers in this domain is to ponder over other annotations, apart from only the handwritten text. A more systematic and theoretical analysis is required for introducing new features that can model a language-independent behavioral writing style of

a person. Further research needs to examine more closely the links between different scripts and then identifying the writer in a multi-lingual setting. It is also observed that a majority of the systems and methods proposed in the literature deal with a noise-free environment. They include documents that are either clean or are segmented well. A robust system must handle documents with noise, such as presence of ruling lines, differential layouts, etc. With the proliferation of deep learning models fast and efficient methods can be developed that can be deployed as end-to-end systems. However, there are limits to how far the idea of automating graphoanalysis can grow. Yet there are certain analyses that deserve efforts: (i) the best discriminatory applied to a selected model (ii) influence of segmentation on performance (iii) finding a specific salient portion of handwriting that distinguishes the best and (iv) Deep architectures that can highlight intra writer variability in due course of life-span for personality assessment. (v) Effects of low-resolution images on writer identification.

II. While extracting annotations in a multi-oriented setting, a major difficulty lies in handling printed text in italics and bold style. They are often misclassified as handwriting. Moreover, there must be provision to consider the graphical content in general. Larger and diverse datasets of annotated documents need to be created. Novel feature sets must be designed for non-controlled environments to differentiate among the annotations and printed text.

III. Reliable extraction of core-region of handwritten words is essential for effective feature extraction. However a larger skew (exceeding 30 degrees) poses a challenge for the core-region extraction methods. Core-region extraction method should also adapt to handle multilingual words. The extracted core regions can also be used for text-line delineation in multi-oriented and multi-layout documents.

In our work we showed that allographic features at grapheme level exhibit discriminative properties when using overlapping windows for writer identification. Future work can focus on developing feature selection strategies to select the best features that can discriminate the graphemes that are part of the same cluster. Features can also be designed to be script specific or script independent, depending on the document being processed.

…