Journal of Universal Computer Science, vol. 14, no. 18 (2008), 2982-2997
submitted: 1/3/08, accepted: 29/9/08, appeared: 1/10/08 © J.UCS

A Language-Independent, Open-Vocabulary System Based on HMMs for Recognition of Ultra Low Resolution Words

Farshideh Einsele, Rolf Ingold
(University of Fribourg, Department of Informatics, Switzerland
farshideh.einsele, rolf.ingold@unifr.ch)

Jean Hennebert
(University of Applied Sciences HES-SO, Business Information Systems, Sierre
Switzerland
jean.hennebert@hevs.ch)

Abstract: In this paper, we introduce and evaluate a system capable of recognizing words extracted from ultra low resolution images such as those frequently embedded on web pages. The design of the system has been driven by the following constraints. First, the system has to recognize small font sizes between 6-12 points where anti-aliasing and resampling filters are applied. Such procedures add noise between adjacent characters in the words and complicate any a priori segmentation of the characters. Second, the system has to be able to recognize any words in an open vocabulary setting, potentially mixing different languages in Latin alphabet. Finally, the training procedure must be automatic, i.e. without requesting to extract, segment and label manually a large set of data. These constraints led us to an architecture based on ergodic HMMs where states are associated to the characters. We also introduce several improvements of the performance increasing the order of the emission probability estimators, including minimum and maximum width constraints on the character models and a training set consisting all possible adjacency cases of Latin characters. The proposed system is evaluated on different font sizes and families, showing good robustness for sizes down to 6 points.

Key Words: ultra low resolution text recognition, web document analysis, HMMs, web image indexation and retrieval
Category: H.2, H.3.7, H.5.4

1 Introduction

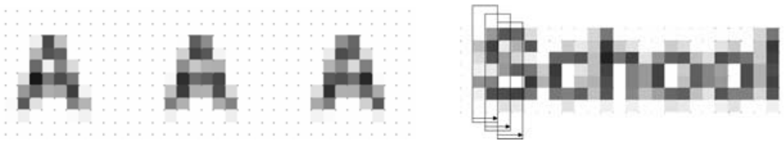
There is no doubt about it: the world wide web has been established as the most ultimative information provider nowadays. Therefore, the investigations in the area of text indexation and information retrieval of web pages has been excessively increased during the recent years. Search engine crawlers have made significant progresses in indexing the HTML plain text. More specifically in the image indexation problem, web pages often contain images with embedded text with important semantical value for information retrieval and indexation [Antonacopoulos et al., 2001a]. Various works report on methods to index the web images. These approaches are either content-based or text-based. The content-based approaches use the image shape, color and texture for search and indexing. They usually compare description of features of a target image with the images contained in their database [Santini, 2002, Scarloff et al., 1997, Benietz et al., 1997]. Content-based indexing is firstly computationally cost intensive when used in a large database and secondly it needs a draft of the searched image to query the database which is not simple nor always available. Text-based indexing analyzes the HTML text associated with these images [Gong et al., 2006]. The analyzed textual description of the image content is stored in a database. Such description can be for example textual information contained in the $< Alt >$ tag or in alternative optional fields, image titles or surrounding text in a HTML page. When the user enters a keyword description of his searched information, this keyword is compared to the description of the stored images in the database. Text-based indexing is obviously computationally less cost intensive because it needs a textual keyword description to search for contextual information. However, both approaches are still insufficient to provide good retrieval quality since the content-based approach does not use textual information embedded in images or in the surrounding HTML text and the text-based approach uses textual annotations in HTML text that is added manually. Textual human annotations are unfortunately mostly poorly provided in HTML pages. Therefore, recognizing text embedded in web images can significantly improve quality of web image indexation and retrieval. One can distinguish between two main categories of text found in web images. The first one corresponds to text visible on scenes shot by cameras. The related area of research which is called "camera based text recognition" tackles this problem [Liang et al., 2005a]. One can find several works in literature reporting promising results to detect and recognize text embedded in low resolution images shot from digital cameras [Liang et al., 2005b, Doermann et al., 2003]. The second category corresponds to bitmap images that are processed using dedicated software such as Photoshop or Fireworks. Such images are generated by web designers to create for example banners, menus, headers, logos etc.. The work that we present in this paper focuses on the recognition of text belonging to this second category. However, we believe that the principles of the approach could also be generalized to camera-based text-recognition.

Such text embedded in web images is often anti-aliased with small font sizes (< 12 pts) and has *ultra* low resolution (between 72 and 90 dpi). An existing approach is to use classical OCR software. However, OCRs are generally built to treat high-resolution (>150 dpi) bi-level images acquired from scanned documents and are therefore not suitable to recognize such text. Several works have been proposing various image enhancement algorithms to transform the text image into a quality that is supported by the existing commercial OCRs [Lopresti and Zhou, 2000, Antonacopoulos and Karatzas, 2004, Perantonis et al., 2003]. While these works have reported noticeable results, the image enhancement is sometimes limited due to very low resolution, to anti-aliasing or to nonhomogeneous text.

Instead of pre-processing the images to feed a classical OCR system, our approach is to use a recognizer specifically trained to recognize such inputs. Our motivation is indeed to reach better accuracy using a recognizer that is specifically trained on inputs including the specificities of such images. To achieve this, we based our system on hidden Markov models (HMMs) that are versatile and powerful statistical tools used in various applications such as in the field of cursive handwriting [Marti and Bunke, 2001, Vinciarelli et al., 2003, Biasdy et al., 2006] or automatic speech recognition [Rabiner and Juang, 1993]. In our previous works, we have been first conducting a preliminary study on isolated characters [Einsele and Ingold, 2005]. The focus was on the understanding of the variabilities of text in ultra low resolution images and on identifying a reliable feature extraction and local character scoring. We have shown that a feature extraction based on moments computation and on multivariate Gaussian density functions leads to robust rsults to recognize isolated characters. Then, in [Einsele et al., 2007b], we extended the approach to recognize full words. Our decision was to use HMMs that present the interesting property to solve the character segmentation and word recognition at the same time. In this approach, one left-right HMM is built for each word where characters are associated to one or more HMM states. A large HMM can be finally built considering the vocabulary taken from a dictionary of 60'000 words, making each word-level HMMs competing against each other. While giving very satisfactory word recognition results, this approach has two drawbacks. First, the recognition is limited to the words available in the dictionary. Some inputs were, indeed, not available in the dictionary, such as inflected forms or proper names, and therefore were not recognized. Second, the memory and cpu usage was still pretty high even when performing several optimization of the HMM topology. A porting of this system on low-end devices such as PDAs would have been difficult to realize.

To overcome these drawbacks, we are proposing here to use an ergodic topology for the HMMs, where all character models are connected to each other. With such a system, the vocabulary size is potentially unlimited while keeping low the usage of system resources. More specifically, we use an ergodic topology based on minimum and maximum constraints which have been obtained automatically from training process. We first measure the impact of number of Gaussian components on a specific font. Then we evaluate the performance of the system when changing font sizes from 6 to 12 points and different font groups. Finally, we perform an error analysis and introduce a so called "balanced training set" to optimize the system.

The remainder of this paper is organized as follows: In Section 2 we list the specifities of text in web images. In section 3 we describe the system used both for training and testing. In section 4 we show the evaluation results and finally



(a) Character level

(b) Word level

Figure 1: (a) Example of anti-aliased, resampled character '`A`' with different grid alignments. (b) Low resolution version of word '`School`' and illustration of the sliding window used for the feature extraction.

we draw conclusions and discuss our future work.

2 Specifities of ultra low resolution, anti-aliased text

The type of inputs our system is treating is illustrated on Fig. 1. These inputs present specificities at the character and at the word level.

Character level: (1) the character has an *ultra* low resolution, usually smaller than 100 dpi with small point sizes frequently between 6 and 12 points, (2) the character has artefacts due to anti-aliasing filters and (3) the same characters can have multiple representations due to the position of the sampling grid.

Word level: As can be observed, there are no spaces as white pixels available to segment characters within the word. Therefore the well-known presegmentation methods [Nagy, 2000] used in classical OCR systems can not be applied anymore in this case. Furthermore, the anti-aliasing noise on both borders of adjacent characters is an additional source of variability.

In our work, we don't treat other artefacts that could be potentially found in text embedded in web images such as color patterns in the character shape or background of the characters, customized spacing between characters, mix characters.

3 System description

We do not address in this work the problem of text detection. Approaches about text detection from web images are reported in [Lopresti and Zhou, 1996, 1997, Antonacopoulos and Karatzas, 2002, 2000, Antonacopoulos et al., 2001b]. Furthermore, we assume that words can be accurately segmented using classical segmentation algorithms like connected components or vertical and horizontal projection profiles. In other words, we are making the assumption that our system receives as input an image including a single word that then need to be recognized by our system.

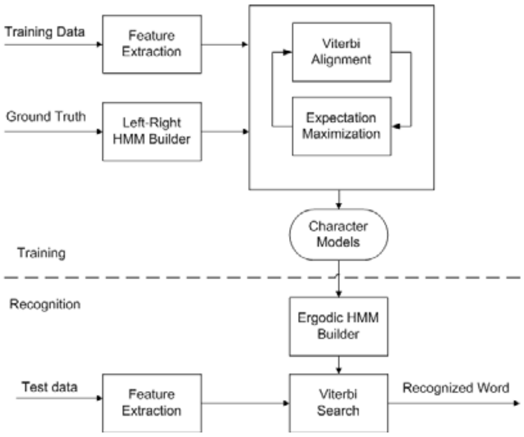


Figure 2: Block diagram of training and recognition

Our system is based on HMMs that are statistical models able to compute the likelihood of an observation sequence given a set of states having transitions between them [Rabiner and Juang, 1993]. Each state is usually associated to a given pattern and a so-called emission probability density function (pdf) is used to model features extracted from this pattern. In our approach, HMM states are associated to character images. The transitions between states are weighted with transition probabilities. Therefore, HMMs model a double stochastic process, one that is tied to the observation of some features (emission probabilities) and one that is tied to the transition between states (transition probabilities). According to this and to the usual simplifying assumptions done with HMMs, the likelihood of a model can be computed as the sum of the product of emission and transition probabilities along all the possible paths.

As illustrated on Fig. 2, our system has two parts: the training part and the recognition part. The training part aims at computing character models by recomposing word-level HMMs based on simple left-right topology iteratively analyzing a large training set of word images. At recognition time, we use an ergodic HMM topology where each character model is connected to each other. More details about the different blocks composing our system are given below.

3.1 Feature extraction

HMMs model ordered sequences of features that are function of a single independent variable. We decided here to compute a left-right ordered sequence of features by sliding an analysis window on top of the word. Therefore the independent variable is, in our case, the x-axis. As *few pixels* are available for each character, we decided to use the *first and second order central moments* as features, since they are translation invariant and can convey sufficient information about a shape even when few pixels are available as in our case. First and second order central moments are described in detail in [H.Bunke and P.Wang, 1997, Gonzalez and Woods, 1992, Jähne, 1997]. We have observed in our previous studies [Einsele and Ingold, 2005, Einsele et al., 2007a] that such features are fairly discriminant for the recognition of ultra low resolution characters embedded in images. As illustrated on Fig. 1(b), we used a 2 pixels length window shifted 1 pixel right. In each sliding window, we compute a feature vector of 8 components including the 6 first and second order central moments, the sum of gray pixel values and an additional feature computed from the difference between the baseline and the y coordinate of the gravity center of each analysis window. This last feature is actually optimistically computed as the baseline is here assumed to be correctly estimated. As output of the feature extraction, a given word image is then transformed into a sequence of feature vectors with 8 components.

3.2 Character model training

The training method is performed directly on words for which a simple left-right HMM is recomposed by gluing together the corresponding character submodels. At training time, a character is modeled with one state where a self-loop transition allows to remain in this model as long as the sliding window is on top of the character. As introduced in [Einsele et al., 2007c], we also use an inter-character model '#' to capture the anti-aliasing noise between the adjacent characters. This model is here treated in the same way as another character model. According to our tests this noisy zone spans 1 to 3 pixels dependent of font family, size and shape. Fig. 3 shows the topology of an HMM recomposed at training time for the word 'cat'.

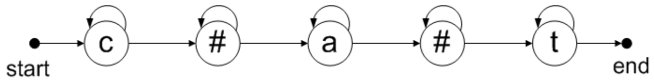


Figure 3: HMM topology for training

The emission probability of each state is computed using continuous mixture of Gaussian components (see for example [Rabiner and Juang, 1993]). The training of the model parameters is performed using an iterative process as follows:

1. **Viterbi alignment**. According to the values of the parameters of the character models, we compute for the whole training set the corresponding alignment between states and feature vectors using the Viterbi algorithm.

2. **Parameter re-estimation**. From the segmentations obtained earlier, we compute new values of the parameters of the character models, i.e. mean vectors, covariance matrices, mixture weights and self-loop probability associated to each state. This re-estimation is also an iterative process that is performed using the classical expectation maximization (EM) process.

Steps 1 and 2 are iteratively repeated until convergence is reached, typically after some iterations. The initial alignment is obtained performing a linear segmentation, assuming that all characters have equal widths. This approach has proven to be efficient provided that the quantity of training samples is large enough. Additionally, we have made the assumption that the components of the feature vector are uncorrelated. This presents the advantage to let the covariance matrix be diagonal and to be more computationally efficient. We have measured that this assumption is actually not critical in terms of accuracy [Einsele et al., 2007c].

3.3 Recognition with ergodic topology

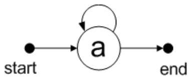


Figure 4: One-state character model

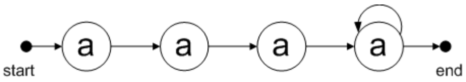


Figure 5: Character model with min width constraint

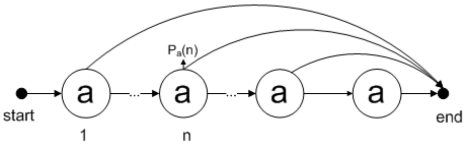


Figure 6: Character model with min-max width constraint

At recognition time, we are proposing here to build a HMM with an ergodic topology at the character level. The topology includes all transitions from one character to the other, therefore allowing to recognize potentially any words in in any language written in Latin alphabet in an open-vocabulary approach. As alternative to the ergodic approach, a topology could also be proposed where a large set of words is used to build a large HMMs where all words are competing in parallel [Einsele et al., 2007b]. While such an approach allows a more precise modeling of a set of words (by including constraints from the lexicon), it has the disadvantage of limiting the recognition capability to a given vocabulary as the Viterbi must search in a HMM made of 60'000 words for the optimal path. In other related recognition domains where the vocabulary of the input is potentially very large, ergodic topologies have also been proposed. We can refer to [El-Yacoubi et al., 2002] where an ergodic HMM system is presented to recognize handwritten street names, to [S. A. Santoshkumar, 2005] for automatic language identification and to [Miyazawa et al., 1994] for speaker verification.

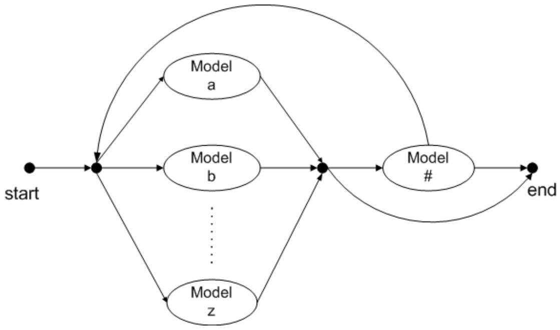


Figure 7: Ergodic topology for testing

Our ergodic topology is illustrated on Fig. 7. As seen on this Figure, the characters are all accessible in parallel and a transition is looping back to all characters from the inter character model '#'. In this Figure, the states represented by black dots are non-emitting states classically used to glue sub-HMMs together. Each character sub-model can take different topology as illustrated on Fig.4-6. The first topology on Fig. 4 is corresponding to the one used at training time. In our previous work [Einsele et al., 2007c], we have experienced that the use of minimum width constraints as expressed in the second topology on Fig. 5 delivers better results as it basically impeach the decoding procedure to leave too early a state giving low local scores. For this work, the minimum width values were inferred from the bounding boxes of each isolated characters. In this paper, we are introducing an extension of this topology that is illustrated on Fig. 6. The values of each transitions leading to the end of the model are corresponding to the probabilities $p_i(n)$ of observing at least n feature vectors in a given character model i. These values are computed during training by inspecting the Viterbi forced alignment on each word. This new width model, while introducing similar minimum width constraints as in model Fig. 5, introduces also a maximum width constraint, expressing the fact that characters have limited width.

For a given test image, we use the Viterbi criterion to determine the best path in this ergodic topology. This path actually defines the recognized sequence of characters composing the word. The Viterbi decoder is also configured to prune out the less probable paths along the recognition process to keep the memory and cpu usage in reasonable ranges.

The transition probabilities going from the submodel '#' back to each character is actually corresponding to the case of equiprobable character sequences, for any pair of characters. We could think of using transition values computed by estimating character bigram frequencies from a real life dictionary. Such a configuration would present the advantage to give lower scores to less probable character sequences, at the cost of making the system dependent to a given language. However, our tests using such character bigram frequencies have shown very little improvements if any in comparison with a simple ergodic topology not based on bigram frequencies. While counter-intuitive, such results can be explained by the fact that transition probabilities have relatively low weight in comparison to emission probabilities. Indeed, emission probabilities are estimated using multivariate Gaussian densities in a high dimensional input space and are therefore leading to quite small values for all feature vectors. Similar observations have been done in the field of speech recognition where it proved beneficial to artificially give more weights to the set of transition probabilities derived from language based constraints. Such approaches, as well as adding character trigram constraints or word bigram or trigram constraitns could potentially improve the system as explained in [Zimmermann and H.Bunke, 2004], but we leave this for future investigations.

4 Experimental results

We focus on single font recognition in this work, which means that each font is modeled and tested independently. Additionally we use the topology of Fig. 6 for all experiments. We use a data base consisting of synthetical word images both for training and test. These word images are produced in high resolutions using the `java.awt.Font` class and are then resampled using Photoshop to our target resolution. Anti-aliasing filters are automatically applied by Photoshop. For testing, an independent set of 3000 unseen word images is generated with the same procedure. We have performed two different evaluation series on the system. Our results are all reported in terms of word recognition rates (WRRs). We experimented with two different training sets. In the first series we were interested in impacts of different system parameters. The second series consists of an error analysis that leads to build a training set to cover the imperfections of the system and benefits from the outcomes of the first experiment series.

4.1 Training with 8'000 word images

A set of nearly 8000 word images is generated by selecting words from a large dictionary. The selection procedure guarantees that each of the 26 characters are represented at least 400 times in the training set. We have investigated the following factors:

1. **Model order**. We investigated the impact of using more complex models by increasing the number of Gaussian components used to compute the emission probabilities. Word recognition rates (WRR) were obtained using the Sans Serif font Verdana, plain, 10 points. As illustrated on Fig. 8, increasing the model order allows to reach better performance thanks to a more precise modeling. No significant gain is observed for models over 64 Gaussians.

2. **Font size**. Using the optimum number of Gaussian components obtained from previous experiments and keeping the same system architecture, we computed the recognition performance for different font sizes going from 12 to 7 points rendered in *plain* style. The objective is here to measure the impact of smaller font sizes on the system performance. Table 1 summarizes the results. As expected, reducing the font size has a negative impact on the recognition performance. Nevertheless, fairly good performance around 91% can still be obtained even for the very small size of 7 points.

Table 1: WRR (%) for ergodic HMM with min-max duration constraints

	Font Size					
	7 pts	8 pts	9 pts	10 pts	11 pts	12 pts
sans serif	91.6	92.0	93.0	93.6	93.7	97.7

4.2 Training with balanced trigrams

We have studied the recognition errors of the above experiments. Typical errors are reported in table 2,where we observe confusions between 'q' and 'd', 'b' and

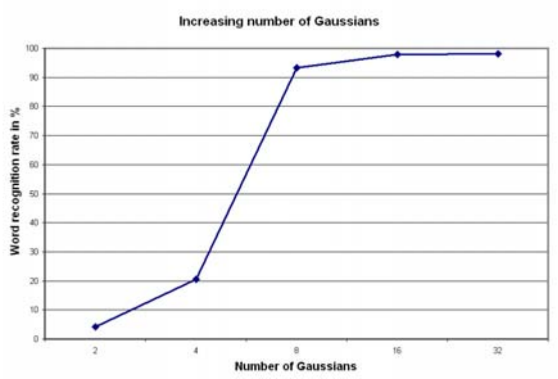


Figure 8: Evolution of WRR when increasing the number of Gaussian components

'p', 'h' and 'n', 'n' and 'r', 'n' and 'h'. Our hypothesis is here that the training set does not contain enough training data for some letters that are frequently confused (for example 'q').

Table 2: Error analysis for ergodic single word recognizer

Genuine word	Recognized word for Font Size					
	12pts	11 pts	10 pts	9 pts	8 pts	7 pts
aldercy	alqlercy	aldercy	aldercy	aldercy	aldercy	alqercy
dakar	qlakar	dakar	qlakar	dakar	qlakar	qdakar
dawn	qlawn	dawn	qlawn	dawn	qlawmh	qdawn
deluded	qleluqled	deluded	qleluded	deluded	qleluded	qdeluqed
enrique	enridue	enrique	enridue	enrique	enridque	enridue
weapons	weapons	weapons	weapons	weapons	weapons	weapbons
bausch	bausch	bausch	bausch	bausch	bausch	bpausch
stann	stann	starhn	starhn	stann	stann	stann
tsang	tsang	tsarhg	tsarhg	tsang	tsang	tsang
stann	stann	starhn	starhn	stann	stann	stann
exiting	exiting	exitirhg	exitirhg	exiting	exiting	exiting

Since the presented ergodic recognizer does not rely on a certain dictionary as shown in [Einsele et al., 2007b] which allows Viterbi decoder to avoid such confusions, we decided to attempt improving the quality of the modelling. In this direction, we enhanced the training set design so that it contains $26 * 26 * 26 =$

Table 3: WRR (%) for open vocabulary system for a serif and a sans serif font with different font styles and sizes using min-max width constraints trained with balanced trigrams

Font family	Font style	Font Size				
		7 pts	8 pts	9 pts	10 pts	12 pts
sans serif	plain	96.9	96.2	98.8	98.8	99.2
	bold	97.1	97.2	98.3	99.1	99.5
	italics	95.7	96.9	97.5	97.9	98.9
	bold+italics	96.8	98.2	98.7	98.9	99.4
serif	plain	93.9	96.1	97.8	98.5	99.5
	bold	94.8	95.1	98.3	98.1	99.9
	italics	95.5	96.1	96.8	97.2	99.3
	bold+italics	96.6	96.6	97.4	98.1	98.8

17′576 trigrams. We tested both serif and sans serif fonts wuith 4 different font styles and for font sizes between 7-12 points. We used a mixture of 64 Gaussian components, as suggested in the previous experiments. The results are listed in table 3. We can see in this table that the recognition rates for sans serif font is indeed considerably improved when comparing to the results from table 1. The improvement of recognition rate is in the range of about 2.5% (for 12 pts) and 6% (for 8,9 pts). When comparing the recognition rates of serif and sans serif fonts, we see that the sans serif has higher recognition rates. Nowadays sans serif fonts like *Verdana*, *Tahoma* and the fonts recently included in Windows Vista like *Calibri*, *Candara* and *Consolas* are mostly used fonts on computer screens. The reason is that generally sans serif fonts show a better legibility than serif fonts for human eyes at ultra low resolution. Interestingly, our system is also better recognizing such fonts. A complete list of Microsoft sans serif fonts can be found in [Microsoft, 2007].

5 Conclusions and future work

We have introduced and evaluated a recognition system capable of recognizing ultra low resolution words extracted from images. The system includes a features extraction module based on sliding windows on which central moments are computed. The modeling part is based on HMMs where each state is associated to a specific character. At training time, the character models are trained automatically by recomposing simple left-right word level HMMs. At recognition time, an ergodic topology is built where all character sub-models are allowed to be followed by any other character model. Minimum and maximum width constraints are also introduced for the character models, simply altering the original topology of each model. We also investigated the impact of building a balanced training set to model equally all potential character sequences. The proposed system has been evaluated on different font families, styles and sizes. From this evaluation, we can conclude on the following advantages of our approach:

- The HMM is able to solve at the same time the segmentation and the recognition of the characters, then avoiding the use of character segmentation procedures that do not apply well on low resolution anti-aliased character sequences.
- The ergodic architecture allows to recognize any words making the system able to work in an open vocabulary manner, potentially on any language supported by this set of characters.
- The robustness of the training convergence of HMMs allows for a fully automated training where the information on character position is not requested.
- The design of the system based on sliding windows, HMMs and multi-Gaussian models allows to apply the same architecture potentially on any font family, style and size.

Potential future works could go in the direction of including linguistic constraints in the system architecture. Including linguistic constraints could be done in a similar manner as in speech recognition systems, using statistical language models including n-gram character constraints. Another possibility could be to keep the n-best recognition hypothesis as output of the ergodic model and to prune out the hypothesis that are unprobable looking in a dictionary. Future works will also be dedicated to the evaluation of the recognition system in a multi-font context and using real-life images extracted from the web.

References

[Antonacopoulos and Karatzas, 2000] Antonacopoulos, A. ; Karatzas, D.: An Anthropocentric Appraoch to Text Extraction from WWW Images. In: *Proc. of the 4th IAPR Workshop on Document Analysis Systems (DAS00).* Rio de Janiro, Brazil, 2000, pp. 515–526

[Antonacopoulos and Karatzas, 2002] Antonacopoulos, A. ; Karatzas, D.: Text Extraction from Web Images Based on Human Perception and Fuzzy Interface. In: *Docuemnt Analysis Systems V.* Princeton, NY, USA, 2002

[Antonacopoulos and Karatzas, 2004] Antonacopoulos, A. ; Karatzas, D.: Text Extraction from Web Images Based on a Split-and-Merge Segmentation Method Using Color Perception. In: *Proc. of the 17th International Conference on Pattern Recognition (ICPR2004).* Cambridge, UK, 2004

[Antonacopoulos et al., 2001a] Antonacopoulos, A. ; Karatzas, D. ; Lopetz, J.O.: Accessing Textual Information Embedded in Internet Images. In: *Proc. of Electronic Imaging, Internet Imaging II.* San Jose, California, USA, 2001

[Antonacopoulos et al., 2001b] Antonacopoulos, A. ; Karatzas, D. ; Lopetz, J.O.: Accessing Textual Information Embedded in Internet Images. In: *Proc. of Electronic Imaging,Internet Imaging II.* San Jose, California, USA, 2001

[Benietz et al., 1997] Benietz, A.B. ; M.Beigi ; S.F.Chang: A Content-based Meta Search Engine for Images. In: *SPIE Proc. of Storage and Retrieval for Image and Video databases*, 1997

[Biasdy et al., 2006] Biasdy, F. ; El-Sana, J. ; Habash, N.: Online Arabic Handwriting Recognition Using Hidden Markov Models. In: *Proc. of Tenth International Workshop on Frontiers in Handwriting Recognition*, 2006

[Doermann et al., 2003] Doermann, D. ; Liang, J. ; Li, H.: Progress in Camera-Based Document Image Analysis. In: *Proc. of Seventh International Conference on Document Analysis and Recognition (ICDAR03)* Ed. 1, 2003, pp. 606

[Einsele et al., 2007a] Einsele, F. ; Hennebert, J. ; Ingold, R.: Towards Identification of Very Low Resolution, Anti-Aliased Characters. In: *Proc. of IEEE International Symposium on Signal Processing and its Applications (ISSPA07).* Sharjah, UAE, 2007

[Einsele and Ingold, 2005] Einsele, F. ; Ingold, R.: A Study of the Variability of Very Low Resolution Characters and the Feasibility of their Discrimination Using Geometrical Features. In: *Proc. of World Academy of Science, Engineering and Technology vol. 6.* Istanbul, Turkey, 2005, pp. 213–217

[Einsele et al., 2007b] Einsele, F. ; Ingold, R. ; Hennebert, J.: Recognition of low resolution word images using HMMs. In: *Advances in Soft Computing 45, Computer Recognition 2.* Springer Verlag, Heidelberg Germany, 2007, pp. 429–437

[Einsele et al., 2007c] Einsele, F. ; Ingold, R. ; Hennebert, J.: Using HMMs to recognize ultra low resolution anti-aliased words. In: *Lecture Notes in Computer Science no. 4815.* Springer Verlag, Heidelberg Germany, 2007

[El-Yacoubi et al., 2002] El-Yacoubi, M. A. ; Gilloux, M. ; Bertille, J. M.: A statistical approach for phrase location and recognition within a text line: an application to street name recognition. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 2*, 2002

[Gong et al., 2006] Gong, Z. ; Hou, L. ; Cheang, C. W.: Web image indexing by using associated texts. In: *Knowledge and information systems, 2006, vol. 10, no2, ISSN 0219-1377.* Faculty of Science and Technology, University of Macau, MACAO, 2006, pp. 243–264

[Gonzalez and Woods, 1992] Gonzalez, R.C. ; Woods, R.E.: *Digital image processing.* Addison Wesley, 1992

[H.Bunke and P.Wang, 1997] H.Bunke ; P.Wang, P. S.: *Handbook of Character Recognition and Document Image Analysis.* World Scientific, 1997

[Jähne, 1997] Jähne, B.: *Practical Handbook on Image Processing for Scientific Applications.* CRC Press, 1997

[Liang et al., 2005a] Liang, J. ; Doermann, D. ; Li, H.: Camera-based analysis of text and documents: a survey. In: *International Journal on Document Analysis and Recognition, vol. 7*, 2005, pp. 84–104

[Liang et al., 2005b] Liang, J. ; Doermann, D. ; Li, H.: Camera-based Analysis of Text and Documents: a Survey. In: *International Journal On Document Analysis and Recognition* Ed. 7, 2005, pp. 84–104

[Lopresti and Zhou, 1996] Lopresti, D. ; Zhou, J.: Document analysis and the World Wide Web. In: *Proc. of IAPR Workshop on Document Analysis Systems.* PA, 1996, pp. 651–669

[Lopresti and Zhou, 1997] Lopresti, D. ; Zhou, J.: Extracting Text from WWW Images. In: *Proc. of the 4th ICDAR*, 1997, pp. 248–252

[Lopresti and Zhou, 2000] Lopresti, D. ; Zhou, J.: Locating and Recognizing Text in WWW Images. In: *Information Retrieval, Volume 2, Numbers 2-3.* Springer Verlag, Heidelberg Germany, 2000, pp. 177–206

[Marti and Bunke, 2001] Marti, U.V. ; Bunke, H.: Using a statistical language model to improve the performance of a HMM-based cursive handwriting recognition system. In: *Int. Journal of Pattern Recognition and Art. intelligence, 15*, 2001, pp. 65–90

[Microsoft, 2007] Microsoft: *Microsoft Typography: Microsoft fonts released in 2007.* http://home.comcast.net/ krieg5208/MS-Fonts.htm. 2007

[Miyazawa et al., 1994] Miyazawa, Y. ; Takami, J.-I. ; Sagayama, S. ; Matsunaga, S.: An All-Phoneme Ergodic HMM for Unsupervised Speaker verification. In: *Proc. Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP94)*, 1994, pp. I–249–252

[Nagy, 2000] Nagy, G.: Twenty Years of Document Image Analysis in PAMI. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (2000), pp. 38–62

[Perantonis et al., 2003] Perantonis, S.J. ; Gatos, B. ; Maragos, V.: A Novel Web Image Processing Algorithm for Text Area Identification that Helps Commercial OCR Engines to Improve their Web Recognition Accuracy. In: *Proc. of the second International Workshop on Web Document Analysis.* Edinburgh, United Kingdom, 2003

[Rabiner and Juang, 1993] Rabiner, L. ; Juang, B.: *Fundamentals Of Speech Recognition.* Prentice Hall, 1993

[S. A. Santoshkumar, 2005] S. A. Santoshkumar, V. R.: Automatic langauage identification using ergodic HMM. In: *Proc. of ICASSP 2005*, 2005, pp. 455–468

[Santini, 2002] Santini, S.: Multimodel search in collections of images and text. In: *Journal of Electronic Imaging, Volume 11, Issue 4*, 2002, pp. 455–468

[Scarloff et al., 1997] Scarloff, S. ; Taycher, T. ; Cascia, M.L.: ImageRover: A content-based Image Browser for the World Wide Web. In: *Proc. of IEEE Workshop on Content-Based Access of Image and Video Librairies (CBAIVL97)*, 1997, pp. 2–9

[Vinciarelli et al., 2003] Vinciarelli, A. ; Bengio, S. ; Bunke, H.: Offline Recognition of Unconstrained Handwritten Texts using HMMs and Statistical Language Models. In: *Transcations of PAMI, IEEE*, 2003, pp. 709–720

[Zimmermann and H.Bunke, 2004] Zimmermann, M. ; H.Bunke: N-Gram Language Models for Offline Handwritten Text Recognition. In: *Proc. of 9th Int. Workshop on Frontiers in Handwriting Recognition (IWFHR04)*, 2004, pp. 203–208