An Interactive Transcription System of Census Records using Word-Spotting based Information Transfer

Joan Mas, Alicia Fornés and Josep Lladós
Computer Vision Center – Dept. Ciències de la Computació. Universitat Autònoma de Barcelona
08193 Cerdanyola del Vallès, Barcelona, Spain
{jmas, afornes, josep}@cvc.uab.es

***Abstract*—This paper presents a system to assist in the transcription of historical handwritten census records in a crowdsourcing platform. Census records have a tabular structured layout. They consist in a sequence of rows with information of homes ordered by street address. For each household snippet in the page, the list of family members is reported. The censuses are recorded in intervals of a few years and the information of individuals in each household is quite stable from a point in time to the next one. This redundancy is used to assist the transcriber, so the redundant information is transferred from the census already transcribed to the next one. Household records are aligned from one year to the next one using the knowledge of the ordering by street address. Given an already transcribed census, a query by string word spotting is applied. Thus, names from the census in time t are used as queries in the corresponding home record in time t+1. Since the search is constrained, the obtained precision-recall values are very high, with an important reduction in the transcription time. The proposed system has been tested in a real citizen-science experience where non expert users transcribe the census data of their home town.**

I. Introduction

In the last decade the increase of mass digitization of historical documents residing in archives and libraries, has carried out the need of (semi)automatic tools for annotation and transcription. This need has increased the exchanges between computer scientists and humanists in the common and growing area of Digital Humanities. Historical documents reflect the identities and contexts of the past, and the access to their contents allows citizens at large to know their individual and collective memory. Historical documents have a great variety, containing hand-written or graphical information in languages and styles evolving over the years. The development of tools for automatic extraction and interpretation of the contents, and cross-linking heterogeneous sources, is still a big challenge. To assist scholars in their digitally-enabled research, and citizens in open access to historical data, new tools and services are being developed for assisted data extraction and annotation. Two key concepts are considered in the design of such systems: the role of the user in the loop, and the use of contextual knowledge.

The inclusion of the user in the transcription and annotation process responds to two demands. First, the assumption that nowadays fully automatic processes are not feasible. Second, the problem is moving to a ”big data” concept. Since many collections are digitized every day, more human power is required to generate accessible data from the contents. This task, that was initially done by experts in archives and libraries, has moved to a popular participation and several crowdsourcing solutions have been proposed. The idea of crowdsourcing is an expression of citizen-science paradigm where the citizens are involved in the systematic collection and analysis of data. It has the advantage of massive data collection but the disadvantage that, since it is based on volunteers and non expert transcribers, if the process is not well designed, it can be annoying and the resulting data can contain many errors. Computer-assisted systems are then gaining relevance, introducing document analysis tools in the platforms. It leans to the second key concept, the use of contextual knowledge. Document sources are highly heterogeneous. General tools for document recognition (e.g. line/word segmentation, writer identification, word recognition or spotting) are not generic enough to keep the performance in different documents of different periods, scripts and topics. Couasnon et al. [1] define two types of contextual knowledge, namely from inside or outside the document page. Inside the document page refers to the terms of the document and their relations (the presence of a term increases the probability of another one). The external knowledge is the correlation and cross-linkage between data of different pages, or collections. The use of contextual knowledge allows to adapt recognition and interpretation tasks to the domain of the processed documents.

In this paper we describe an assisted crowdsourcing system for transcribing historical census records where the modelling of internal and external context significatively increases the performance of human transcribers. Demographic data analytics tools allow scholars to generate genealogies, to establish individual and family lifespans, to study migrations, and to spatially locate family networks, among other tools. Census records have a tabular structure collecting information of individuals in each household at a certain point in time such as names of family members, ages, place of birth, family relationship, occupation, value of their home and belongings, etc. The information is ordered by street address. Census are recorded in cities at regular points in time. This structure and procedure is quite common world-wide. This is a key observation in this work. The structure of census records

allows to model the above described internal and external context.

Therefore, in this work we propose a system for data extraction of census records taking advantage of the knowledge of the structure. The internal context is defined in terms of the layout of a page (homes ordered by street address). The external context is defined in terms of the principle that the home information at time t is highly coincident with the information of the census at time $t-1$. Hence, once the records of a census have been transcribed by human operators, this information is "dragged" to the corresponding records of the next one. Figure 1 shows the architecture of the proposed system. First, the logical layout of a page is extracted. It results in a list of regions of interests corresponding to homes (roughly correspond to perceptual table rows). The user is requested to label these regions with the street address. We assume that the census at time t-1 has been already transcribed, so the information corresponding to family names of the same home address are searched in the image region using a query by string word spotting process. There is a high probability of finding the same family names, or small variations, in the same record 3-5 years later. In the transcription process, the user is assisted by the system, that suggests the names of the record being processed, transferring this information from the record at time t-1. The proposed system is used in a real scenario. The census between 1828 and 1955 (19 census) of the Catalan city of Sant Feliu de Llobregat are transcribed by citizens of the municipality. To illustrate the effectiveness of the concept of information transfer described in this paper, experiments using the census records of the years 1881 and 1886 have been done. The fist year is used as reference, and the second one is semi-automatically transcribed with the linked information. We will experimentally show that an important save of time is achieved in the process.

This paper is organized as follows. In section II we review some systems for data transcription of historical documents. Section III describes our system and the key contributions. In section IV we describe the experimental framework. Finally section V draws the conclusions.

II. Previous works on transcription tools

We focus on digitized historical demographical manuscripts as an example of one of the central interests of digital humanities. Several projects have dedicated important resources to digitize and extract the information of these documents, such as FamilySearch [2], the Ancestry.com World Archives Project [3] [4], or the Mormon Migration project [5].

Concerning transcription and annotation, several ground-truthing tools have been developed, such as DEBORA [6], ALETHEIA [7], GiDoc [8] or Pixlabeler [9]. In case the amount of documents to be transcribed is significantly large, a crowdsourcing paradigm [10], [11] is typically followed. Thus, the amount of work is distributed into small and simple tasks (e.g. transcribe few pages, lines or words) among the community of transcribers, speeding up the process. Some of the most popular are Transkribus [12], Shared Canvas [13] for annotating Digital Facsimiles, Transcribe Bentham [14], the Digital Vercelli Book [15] or DigiPal [16]. In this sense, it is worth to mention the *I ANNOTATE* [17] initiative for fostering a worldwide community for annotation.

Referring crowdsourcing platforms for demographic documents, one can find the Ancestry.com Keying Tool [18] the Civil War Diaries and Letters Transcription tool [19], [10], the Citizen Archivist Dashboard [20], or the Transcription Project tool [21], and the marriages records tool [22] for the 5CofM project [23].

The collaborative transcription is even faster when using intuitive multi-touch devices [24]. However, in most of these applications the preprocessing and segmentation of text blocks is required before the annotation step starts. The extreme cases are the ones that involve segmentation of the document images into individual words, which is the case of reCaptcha [25] or Digitalkoot [26].

In parallel, document image analysis techniques, such as handwriting recognition [27] and word spotting [28] are also investigated for helping in the massive extraction of information from document images. Since a complete automatic transcription is unfeasible, the alternative is to follow a semi-assisted process, such as the multi-modal interactive transcription tools [29], [30], [31] developed for the TranScriptorium project [32].

III. System Description

The proposed system aims to help the user in the transcription of a new document by taking advantage from the already transcribed census documents of previous ages. The architecture of the system is divided in three main stages: Analysis and extraction of the layout of the document, recognition of layout identifiers and retrieval and spotting of previous information. Figure 1 illustrates the process. Given an image, its layout is segmented in regions corresponding to the different homes in the document. Then, we label them according to the street and number. Afterwards, the information of the different inhabitants in the home is obtained from the database of transcribed census of previous years, indexing by the street address. Finally, the obtained family names are used as search terms in a query by string word spotting performed on the region of the document belonging to that specific home. When the system is not able to match a single instance of the queried names, a list of possibilities is shown to the user, so he/she can select the right one or discard the suggestions and type the correct transcription. Figure 2 shows the interface of this step. In the following subsections, the different steps are further described.

A. Layout Analysis for Home Segmentation

This module is driven to extract the different homes in a document. The steps are the following. First, we binarize the document using a method based on a local neighborhood [33]. Then, we perform a vertical projection to extract the different columns of the table. We search the column that contains the street number and extract the different candidates to street

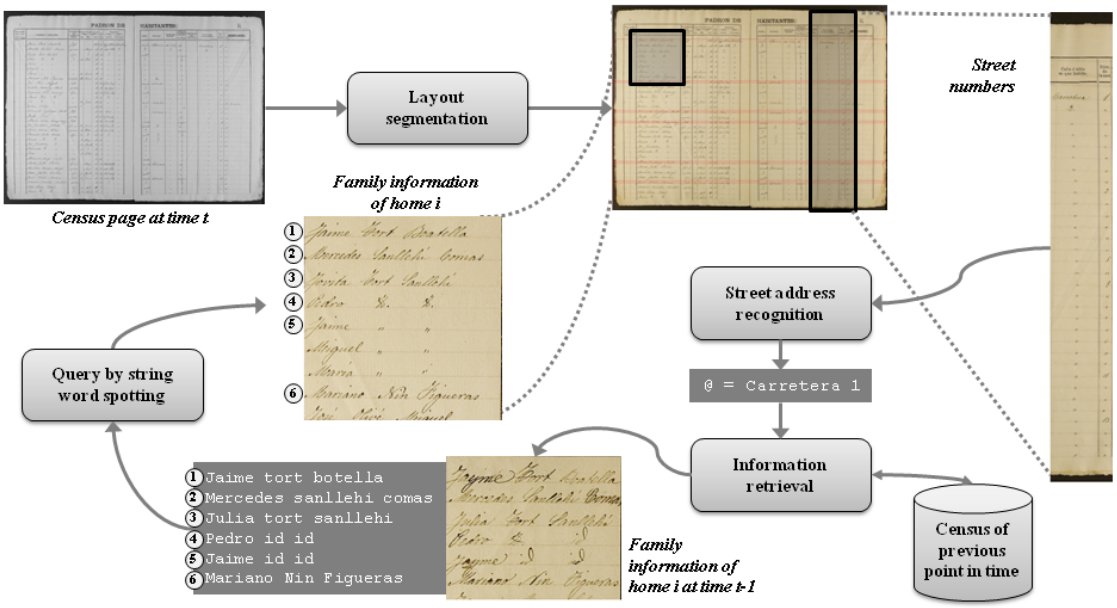


Fig. 1. Architecture of the proposed system.

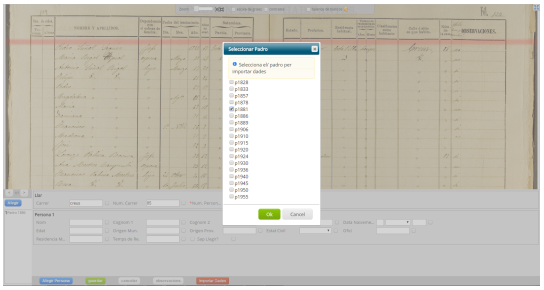


Fig. 2. Interface of the proposed system when several family names are suggested after a word spotting to the previous census.

numbers. Some heuristics are applied in order to reject those candidates that are not feasible to be a street number. In addition, we have defined a model to detect the colon and the word "id", since it is used to denote the same street number than the previous one.

B. Street Number Recognition

This module allows to identify which is the concrete street name and number to label a candidate home. Once we have identified the candidate numbers in the previous step, we recognize them by using the combination of 5 different Multi-Dimensional Long Short-term neural networks (MDLSTM) [34]. There is no feature extraction since the raw pixels of the image are used as features for each piece at the input layer of the neural network. The best solutions are shown to the user who selects the good one or corrects it in case the correct one is not shown. The networks have been trained on 10.000 images of sequences of digits, so we do not need to segment the street numbers into individual digits.

C. Information Retrieval and Spotting-based Transfer

This module is devoted to put into correspondence the information of the current document with the information contained in a previous transcribed census. Basically, the goal is to link each of the images of text lines corresponding to the inhabitants in the current home (e.g. a census from 1884) with the names that appear in the previous home (e.g. a census from 1882), which have been already transcribed and stored in the database.

For this purpose, we take each one of the names and surnames in that home (from the previous census) and search them into the image using a query-by-string word spotting strategy based on attributes [35]. The query-by-string word spotting method embeds in a common space the feature vector from the word images (described with Dense SIFT features), and the feature vectors from text strings (described with Pyramid of Histogram Of Characters) by using Canonical Correlation Analysis. In this way, the input text string is projected in a common space in order to find its corresponding word image. Afterwards, the Hungarian method is used to match regions in the image corresponding to the line with words in the previous census information. The goal is to find individuals, which means that the name and surnames of an individual have to be found in the same text line.

We use a word spotting methodology instead of a Handwriting Text Recognition (HTR) method because we have the information of the previous census, and consequently, the problem can be reduced to match the names in the current document image with the names stored in the database. In addition, we aim to check the existence of these inhabitants in the previous census, instead of transcribing these words. Another reason is that, contrary to most HTR systems, we avoid to define any language model.

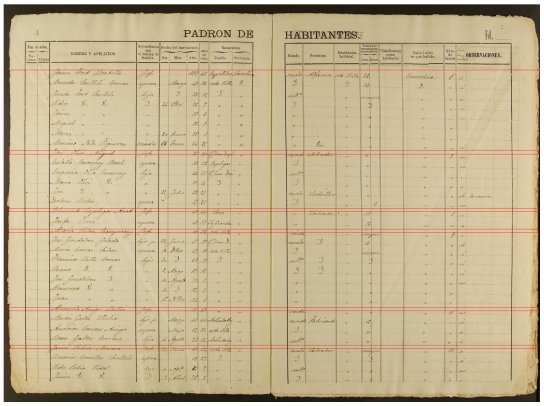


Fig. 3. Example of a document from the census of the Catalan city of Sant Feliu de Llobregat.

IV. Experimental Evaluation

The proposed system has been experimentally evaluated in a real use case consisting in a transcription campaign of the census records of a Catalan city, with the participation of citizens. It is part of a project for the digitalization and valorization of demographic sources from the past. To evaluate the performance of the different algorithms proposed in the system, three different experiments have been designed: home segmentation (layout analysis module), street number recognition, and information retrieval and spotting.

A. Experimental Framework and Dataset

A crowdsourcing experience involving citizens has been carried out in the Catalan city of Sant Feliu de Llobregat. Citizens of the city transcribe the census of the period of 1828 to 1955. The collection consists of census collected at 19 points in time, with more than 30,000 images in total. The number of registered inhabitants in this period ranges between 1,500 and 7,000. Figure 3 shows an example of a document. The information of homes is written ordered by street address. The third column has the names of the family members. The third and second columns from the right contain the street name and number respectively. As it has been explained previously, since the structure of the census is stable from one point in time to the next one, the correspondence between the snippets of homes allows to transfer the information already transcribed to the new records. Due to socio-economic reasons, there are some years when the information contained in the records are more stable. These years are used as references for transferring information. Two candidate years are 1881 and 1940. In this paper, we have used the census of 1881 as reference, so it has been manually transcribed. The census of the year 1886 has been used as target where transfer the information. In the next sections the experiments on the different steps are reported.

B. Layout Analysis for Home Segmentation

To evaluate the home segmentation step we have selected 191 documents (97 from the census of 1881 and 94 from the one of 1886). The precision and recall metrics on the number of homes detected and segmented have been computed for the evaluation. Table I shows the performance. As can be seen the method performs a good segmentation obtaining precisions of 90% and recalls of 88%. The most common errors are split or merge of regions, or the unability to detect homes because the reference street number is missing. These errors are due to the confusion of quotation marks or the word *id* by street numbers when actually it refers to people of the same home, or when houses differ in the floor number.

TABLE I
Results on home segmentation for the census documents of 1881 and 1886

Census	*Precision*	*Recall*
1881	0.8995	0.8810
1886	0.9289	0.8928

C. Street Number Recognition

For this experiment we have selected 11 different documents with a total of 66 home regions. After segmenting the regions of interest containing the street number, a slant correction preprocess has been performed. The numbers are recognized with the MDLSTM neural network classifier. This unsupervised approach does not resulted in a good performance (a score of 48%) due to confusion of some digits (e.g. 3 and 9). To increase the recognition rate we have made use of the order information of street numbers. Thus, following an ordinal classification strategy the performance increased to 85%.

D. Information Retrieval and Word Spotting

In this experiment we have evaluated the linkage and transfer performance between the census of 1881 and the snippets of family names of the census of 1886. For the sake of understandability we will refer as *word strings* the transcribed names of the census of 1881, and as *word images* to the snippets of cropped images of the census of 1886. Thus, given a word image to be recognized while the document is transcribed, it is searched in the database of word strings of the previous census already transcribed. This search is done with the query by string word spotting method. The correct matched word string is "transferred" to the current census record as transcription of the word image. Since this method requires a training process, the model has been trained with 247 words extracted from a previously transcribed census.

We implemented two scenarios. Given a household image of the census being transcribed, the first scenario performs a one to many linkage, i.e. it does not search for the corresponding record in the transcribed census but all the transcribed names are compared. In the second scenario, referred as one-to-one linkage, only the information of the corresponding home is used from the previous census.

1) One-to-many Linkage: We used 301 words (strings) corresponding to 133 persons (a person has one name and one or two surnames) from the transcribed records of 1881. On the other hand, we used 2313 query snippets corresponding

TABLE II
$P@k$ FOR k = 1,2,3,4,5 AND 10

k	*Prec(k) (# words)*
1	**0.4153 (125)**
2	**0.5083 (153)**
3	**0.5382 (162)**
4	**0.565 (170)**
5	**0.598 (180)**
10	**0.688 (207)**
Total Words	**301**

to cropped word images after the homes of the 1886 census were segmented. Given a query word image w, a ranked list is obtained computing its distance to the word strings of the previous record. We consider that only one matching is correct, hence the goodness of the retrieval is defined by the score $Sc(w) = 1/Pos(w)$ where Pos is the position of the correct retrieval. With this metric, we obtained an average score of 0.50 which corresponds to the average 2nd position in the retrieved ranked lists. We have counted the number of people correctly retrieved, considering as correct when the sequence of name and surnames is found within the top 10 positions. This number is **62** which corresponds to a **46.6%**. In table II we present the precision at k ($P@k$) for the top five and tenth positions. In the application, the users are provided with the top ten retrieved words. consider that validating more than ten requires more effort that directly write the transcription.

2) One-to-one Linkage: For the second scenario we have selected the images of three streets that contain of 104 homes and 504 people. Once homes are segmented and labelled after recognizing the street address, each home is linked to the corresponding transcribed record in the previous census. Word images are compared to the word strings of the linked record using the query by string method. As in the previous scenario, words in the same line likely corresponding to names and surnames of the same inhabitant are searched together.

The system is not able to retrieve the information in the following cases:

- Family members of a home have moved to another house.
- New born or dead members in the period between the two census.
- One of the members have been married.
- Some service personnel has changed or has been hired.

We have defined a metric at record level based in the Levenshtein distance in terms of the number of edit operations of word recognitions, i.e. insertions correspond to new inhabitants, deletions are inhabitants not found, and substitutions are erroneous correspondences. Thus the metric estimates the effort that the user must do in the transcription. Thus, given a record r, the score is defined as follows:

$$Sc(r) = 1 - \frac{\#Operations}{\#Members} \quad (1)$$

where *#Operations* is the number of editions and *#Members* is the number of inhabitants (lines) in this record. For the current configuration, we have obtained an score of **0.7024**.

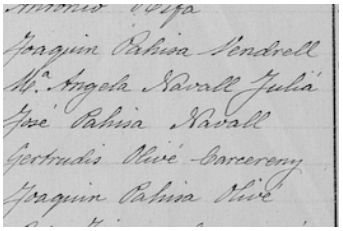
TABLE III
NUMBER OF OPERATIONS NEEDED IN THE EXPERIMENT

Operations	*Value*
Total	**150**
Substitutions	**25**
Insertions	**125**

The true positive rate (TPR) is of 0.9346. It results in a reduction of 70% in the transcription time. In our transcription experience a user took 30 sec. to enter a person. Thus, for 504 inhabitants the effort of 4.2 hours is reduced to 1.25 hours.

Table III reports the number of operations needed in the experiment. The higher amount corresponds to insertions. Semantically, it is due to the birth of new members in a home or the move from one home to another.

Figure 4 presents a qualitative example of the search for the family member *joaquin pahisa vendrell*. We can see the image corresponding to the record, in the left, and the similarities of the query by string word spotting comparison. The first and the fifth lines have the higher similarities because they only differ in the last word. Figure 5 shows another example when the query is *juan marti canameras*. In this case, the similarity is 0 in the two cases, hence the query is discarded and the user is requested to type the transcription.

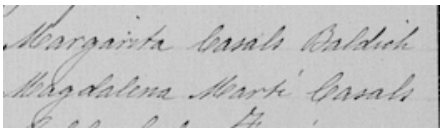


line 1	0.4883
line 2	0.2073
line 3	0.3258
line 4	0.0
line 5	0.4174

(a) (b)

Fig. 4. Example of retrieval (a) Image region corresponding to the household. (b) Similarities regarding the query *joaquin pahisa vendrell*.

Comparing the two scenarios, the latter offers a better performance (**23%** approximately). Nevertheless, the first procedure allows to retrieve names and surnames of persons that are not present in the linked record of the previous census, although all the information of the person is not complete. To conclude, the combination of the both methods would help to increase the percentage of information transcribed since we can recover information from people moved from one home to another or from homes that are not in the previous census.



line 1	0
line 2	0

(a) (b)

Fig. 5. Example of retrieval (a) Image region corresponding to the household. (b) Similarities regarding the query *juan marti canameras*.

V. Conclusions

In this paper we have presented a system to assist the user in the transcription of historical census documents. The temporal continuity of one census regarding the one registered a few years before is used to link the records corresponding to homes and transfer the information. The knowledge about the tabular structure of a census is used to segment the home regions of the image. Each region is labelled with the street number recognized with an ordinal classification strategy based on a MDLSTM neural network classifier. The system uses a previous record already transcribed as reference. Thus, the inhabitants names are recognized using a query by string method, and transferred to the record of the census being transcribed. In this process, the user is kept in the loop. Thus, the record linkage and information transfer procedure is integrated in a crowdsourcing platform to assist the human transcriber. The proposed system has been used in a real experience where a set of citizens collaborate in the transcription of population censuses of their city. With the proposed strategy of using the temporal context, the transcription time is dramatically reduced over a 70%. The confidence of the recognition high with a 93.46% of true positive rate. The cases in which the names are not linked are generally due to the contents but not to errors of the recognizer. Thus changes are due to new born or dead people, or movements to different homes. The linkage of census records is not only useful for transcription but gives added value to the demographic data. It is a tool for scholars in humanities for life course and genealogy analysis.

Acknowledgement

The research leading to these results has received funding from RecerCaixa, the European Research Council Advanced Grant (ERC-2010-AdG 20100407: 269796-5CofM) and the Spanish grants TIN2012-37475-C02-02 and RYC-2014-16831. We thank the volunteers of Sant Feliu that transcribed the census documents and the Arxiu Comarcal del Baix Llobregat for the images.

References

[1] B. Coasnon, "Dmos, a generic document recognition method: application to table structure analysis in a general and in a specific way," *International Journal of Document Analysis and Recognition (IJDAR)*, vol. 8, no. 2-3, pp. 111–122, 2006. [Online]. Available: http://dx.doi.org/10.1007/s10032-005-0148-5

[2] "Familysearch." [Online]. Available: https://familysearch.org/

[3] "Ancestry.com world archives project." [Online]. Available: http://community.ancestry.com/awap

[4] A. G. Noll, "Crowdsourcing transcriptions of archival materials," in *Graduate History Conference*, march 2013, pp. 1–33. [Online]. Available: http://scholarworks.umb.edu/ghc/2013/panel6/4/

[5] "Mormon migration project." [Online]. Available: http://mormonmigration.lib.byu.edu/

[6] F. Le Bourgeois and H. Emptoz, "Debora: Digital access to books of the renaissance," *International Journal of Document Analysis and Recognition (IJDAR)*, vol. 9, no. 2-4, pp. 193–221, 2007.

[7] C. Clausner, S. Pletschacher, and A. Antonacopoulos, "Aletheia-an advanced document layout and text ground-truthing system for production environments," in *International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 2011, pp. 48–52.

[8] "Gidoc tool." [Online]. Available: https://www.prhlt.upv.es/page/projects/multimodal/idoc/gidoc

[9] E. Saund, J. Lin, and P. Sarkar, "Pixlabeler: User interface for pixel-level labeling of elements in document images," in *10th International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 2009, pp. 646–650.

[10] S. Averkamp and M. Butler, "The care and feeding of a crowd," in *Code4Lib Conference*, February 2013. [Online]. Available: http://code4lib.org/conference/2013/averkamp-butler

[11] M.-C. Yuen, I. King, and K.-S. Leung, "A survey of crowdsourcing systems," in *IEEE third International Conference on Privacy, security, risk and trust (PASSAT), and IEEE third International Conference on Social Computing (Socialcom)*. IEEE, 2011, pp. 766–773.

[12] "Transkribus." [Online]. Available: https://transkribus.eu/Transkribus/

[13] R. Sanderson, "Shared canvas: Digital facsimiles via distributed annotation," in *I Annotate - Annoto Ergo Sum*, San Francisco, 2013.

[14] "Transcribe bentham." [Online]. Available: http://www.transcribe-bentham.da.ulcc.ac.uk/

[15] "Digital vercelli book." [Online]. Available: http://vbd.humnet.unipi.it/beta/index.html#104v

[16] "Digipal." [Online]. Available: http://www.digipal.eu/digipal/page/

[17] "I annotate." [Online]. Available: http://iannotate.org/

[18] "Ancestry.com keying tool." [Online]. Available: http://www.ancestry.com/wiki/index.php?title=Keying_tool_how-tos

[19] "Civil war diaries and letters transcription." [Online]. Available: http://diyhistory.lib.uiowa.edu/

[20] "Citizen archivist dashboard." [Online]. Available: http://www.archives.gov/citizen-archivist/

[21] "Transcription project tool." [Online]. Available: http://transcribe.archives.gov/

[22] A. Fornés, J. Lladós, J. Mas, J. Pujades, and A. Cabré, "A bimodal crowdsourcing platform for demographic historical manuscripts," in *Digital Access to Textual Cultural Heritage Conference (DATeCH)*, 2014, pp. 103–108.

[23] "5cofm. five centuries of marriages project." [Online]. Available: http://dag.cvc.uab.es/infoesposalles/

[24] A. Amato, A. Sappa, A. Fornés, F. Lumbreras, and J. Lladós, "Divide and conquer: Atomizing and parallelizing a task in a mobile crowdsourcing platform," in *2nd International ACM Workshop on Crowdsourcing for Multimedia (CrowdMM)*, 2013, pp. 21–22.

[25] "recaptcha." [Online]. Available: http://www.google.com/recaptcha

[26] "Digitalkoot." [Online]. Available: http://www.digitalkoot.fi/index_en.html

[27] T. Plotz and G. Fink, "Markov models for offline handwriting recognition: a survey," *International Journal on Document Analysis and Recognition (IJDAR)*, vol. 12, no. 4, pp. 269–298, 2009.

[28] V. Frinken, A. Fischer, R. Manmatha, and H. Bunke, "A novel word spotting method based on recurrent neural networks," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 2, pp. 211–224, 2012.

[29] R. McNicholl and T. Miles-Board, "Transcriptorium : computer-aided, crowd-sourced transcription of handwritten text (for repositories?)," in *10th International Conference on Open Repositories (OR2015)*, 2015, pp. 21–22.

[30] B. Gatos, G. Louloudis, T. Causer, K. Grint, V. Romero, J. A. Sanchez, A. H. Toselli, and E. Vidal, "Ground-truth production in the transcriptorium project," in *Document Analysis Systems (DAS), 2014 11th IAPR International Workshop on*. IEEE, 2014, pp. 237–241.

[31] D. Martín-Albo, V. Romero, and E. Vidal, "Escritoire: A multi-touch desk with e-pen input for capture, management and multimodal interactive transcription of handwritten documents," in *Pattern Recognition and Image Analysis*. Springer, 2015, pp. 471–478.

[32] "Transcriptorium project." [Online]. Available: http://transcriptorium.eu/

[33] D. Bradley and G. Roth, "Adaptive thresholding using the integral image," *Journal of graphics, gpu, and game tools*, vol. 12, no. 2, pp. 13–21, 2007.

[34] A. Graves and J. Schmidhuber, "Offline handwriting recognition with multidimensional recurrent neural networks," in *Advances in Neural Information Processing Systems*, 2009, pp. 545–552.

[35] J. Almazan, A. Gordo, A. Fornes, and E. Valveny, "Word spotting and recognition with embedded attributes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 12, pp. 2552–2566, 2014.