

The MUMTDB dataset for evaluating simultaneous composition of structured documents in a multi-user and multi-touch environment

Zhaoxin Chen, Eric Anquetil, Harold Mouchère, Christian Viard-Gaudin

► To cite this version:

Zhaoxin Chen, Eric Anquetil, Harold Mouchère, Christian Viard-Gaudin. The MUMTDB dataset for evaluating simultaneous composition of structured documents in a multi-user and multi-touch environment. 15th International Conference on Frontiers in Handwriting Recognition, Oct 2016, Shenzhen, China. <hal-01398762>

HAL Id: hal-01398762

<https://hal.archives-ouvertes.fr/hal-01398762>

Submitted on 17 Nov 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The MUMTDB dataset for evaluating simultaneous composition of structured documents in a multi-user and multi-touch environment

Zhaoxin Chen and Eric Anquetil
INSA de Rennes
Rennes, France 35043
Email: zhaoxin.chen@irisa.fr
eric.anquetil@irisa.fr

Harold Mouchère and Christian Viard-Gaudin
LUNAM, University of Nantes, IRCCYN
Nantes, France 44300
Email: harold.mouchere@univ-nantes.fr
christian.viard-gaudin@univ-nantes.fr

Abstract—We propose in this paper a new online Multi-User Multi-Touch handwritten diagram DataBase (MUMTDB) for evaluating recognition systems under the multi-user situation. The data is collected according to two predefined mind map scenarios which contains 9 classes of graphical symbols. Each scenario is completed by involving two users at the same time. Since the users are given freedom to draw the symbols as they want, the dataset contains a diversity of multi-stroke and even multi-touch symbols. It allows addressing new challenging problems regarding the recognition of simultaneous composition of structured documents. The dataset is freely available on-line.

Keywords—Multiple users; multi-touch gesture; handwritten diagram; stroke grouping;

I. INTRODUCTION

A large multi-touch display allows multiple users to simultaneously interact in the same context and work together. Indeed, many researches and commercial products propose tangible interfaces which support simultaneous participation of multiple users. To the best of our knowledge, there are few research focusing on the freely-drawn sketch case for multiple users.

In this paper we present a multi-user handwritten diagram dataset. The handwriting data is collected based on two predefined mind map scenarios which contains 9 different classes of graphical symbols. A key feature is that each scenario is completed by involving two users at the same time (as shown in Fig.1). Since there is no constrain no the users for composing the symbols, the dataset contains a diversity of multi-stroke and even multi-touch symbols. This dataset is built for two main reasons: firstly, to analyze the user's behavior and evaluate the recognition strategy for multi-stroke and multi-touch symbols in the multi-user environment; secondly, to explore a dynamic strokes grouping and recognition method in order to give a real-time feedback to the users.

A user can draw two kinds of gestures on a multi-touch screen : symbol gestures to add new content in the document and command gestures to interact with the document and software. In a previous work [1] we have proposed a solution

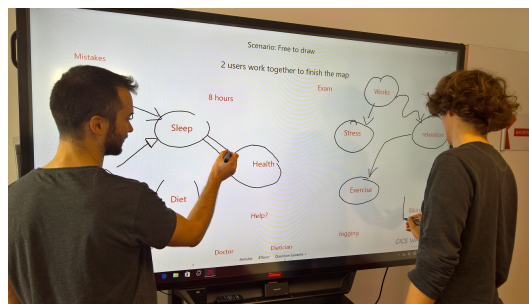


Figure 1. The diagram data acquisition procedure on a 80" touch screen. Two users are drawing the diagram together using styluses.

to deal with early recognition of command gestures but with a single user. This new dataset will allow the gesture recognition community to address the problem of grouping the strokes from different writers when each of them are able to interact with several fingers at the same time at anyplace without constraint. Our mid-term objectif is to be able to design a real-time feedback recognition system which avoid any constraints for users. The users should be able to draw the diagrams as on a real whiteboard.

One of the most difficult challenges in multi-user sketch is clustering the strokes into correct groups to form valid symbols. For a real-time feedback application, a more natural way is to automatically detect the terminal stroke of a symbol and recognize the symbol within an acceptable time delay T_d . Fig. 2 illustrates how should be a real-time recognition system : the users draw strokes which are accumulated in a temporally local context. Then for each new stroke, the system should try to recognize one or more gestures using these strokes, each time a gesture is recognized its strokes are removed from the context. We can see that the time becomes an important parameter for the interpretation of the strokes.

In section II we describe some existing solutions for stroke grouping examining different contexts. Then in section III we present our Multi-User diagram database and illustrate the difficulties of the recognition task. Finally the last section

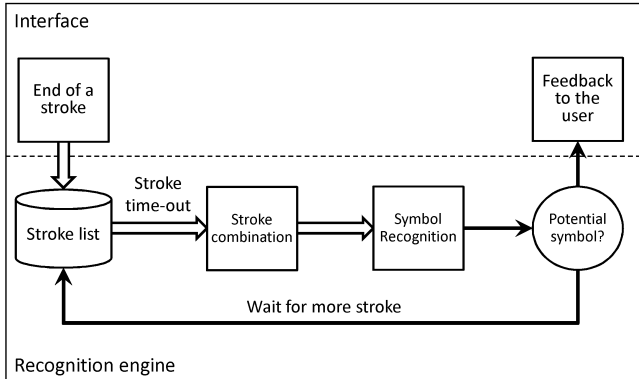


Figure 2. The framework of a real-time grouping and recognition system.

concludes with suggestions for solving the problem.

II. RELATED WORK

The handwritten sketch and graphic symbol recognition problem in online document has been widely studied in recent years. However, there is few works on sketch and symbol recognition in the specific context of multi-touch and multi-user. The need for a benchmark database supporting the development of the systems is a central issue for the community. We will firstly summarize some basic principals and database that have been used to address on-line sketch recognition.

There has been plenty of researches to design very powerful on-line recognition systems for isolated hand-drawn symbols. But for the diagram recognition, where a stroke can have various different interpretation according to its context, the task of grouping strokes brings serious problems for the analysis process. A simplest *mono-stroke* solution is to require users to draw each symbol with only one stroke and limit the system to be used by single user [2] [3]. A typical mono-stroke graphical symbol dataset is ILGDB [4]. Furthermore, Awal et al. [5] propose a *multi-stroke* flowchart database containing both graphical symbols and text, which generally matches the way people naturally draw. To solve the stroke grouping problem, they generate most of possible strokes segmentation filtered by some geometric constraints. The final segmentation is the one which maximize the probability of each symbol recognized by a classifier. [6], [7] also present their text/non-text separation approaches and structure analysis on this database. Meanwhile, it is also possible for a user using *multi-touch* manner (i.e. using two or more fingers at the same time to produce a symbol) to compose a graphical symbol such as 'double line' or 'double bracket'. In our previous work we [1] analyzed the characteristic of *multi-touch* graphical symbols and proposed a *multi-touch* handwritten isolated gesture database. However, all the previous works are limited to a mono-user context.

As defined in [8], we can distinguish two kinds of approaches to analyze a handwritten diagram: the traditional

one is based on *lazy interpretation* [9]. Such systems analyze the user strokes when the diagram is explicitly finished. The analysis process globally interprets all the strokes including stroke segmentation and structure analysis. In [7] the recognition process is guided by a grammar based syntactic analysis which parses the full document structure. Bresler et al. [10] estimate a general distance threshold based on the two closest points from two strokes to determine if they are enough spatially close to be a symbol candidate. Then the final decision is seen as a max-sum optimization considering the symbol probabilities and their relations. But since the lazy interpretation analyze the diagram regardless user's creative phase during the composition, it is not possible to give a real-time feedback to the user, which reduces its usability.

Instead of analyzing the global structure of the diagram as these works, an on-line multi-user freely-drawn system is supposed to give a real-time analyzing including stroke grouping and recognizing. This is defined as *eager interpretation* [8] which tries to understand the structure of the diagram as well as its elements during its composition, more precisely after each input stroke. This offers many benefits from a human-computer interaction point of view in comparison with lazy interpretation. Peterson et al. [11] present a two step solution: first isolated stroke are classified and then a binary stroke grouping classifier (based one one temporal feature and 12 spatial features between two strokes) is used to incrementally group the stroke. Mace et al. [8] parse the strokes taking into account the global vision of the current state of the document with a local vision of a symbol. The associated parser exploits both the formalized knowledge and the temporal context to interpret each input stroke.

In our context, we go one step further to open a new frontier for the *multi-user* handwriting recognition. It is the first database of online handwritten diagrams containing both *multi-touch* and *multi-stroke* symbols from *multi-user*. This context has a more complex spatial and temporal relations between strokes which make it more difficult for analysis.

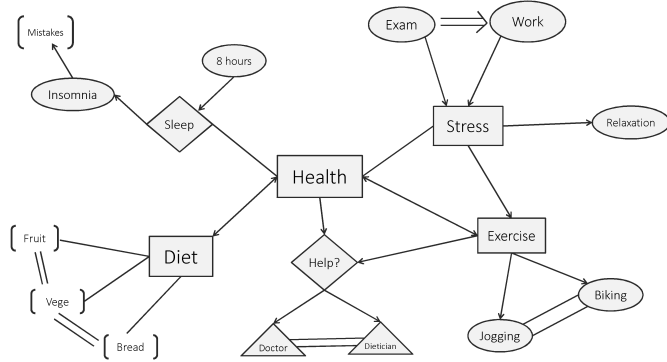
III. MULTI-USER DIAGRAM DATABASE

Since there are few works on multi-user handwritten analyzing, it is essential to begin with a representative dataset. We will present the procedure of data acquisition and the multi-user features of the diagrams in this section.

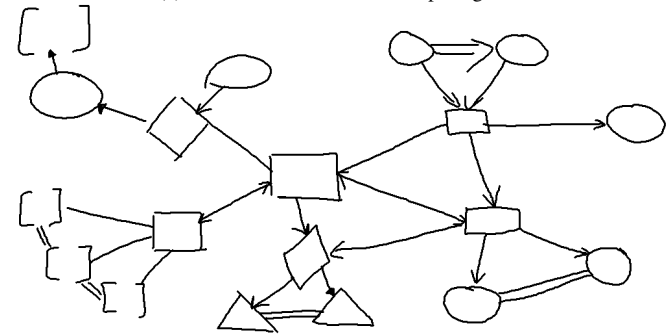
A. Diagram acquisition

We propose the mind map diagram as the multi-user data acquisition scenarios. Each time two users were asked to stand in front of a touch screen side by side and complete a mind map diagram together. An example of a mind map diagram is shown in Fig. 3 (a). As a first step towards a more complex dataset, we focus more on the graphical symbols rather than the text. Therefore, the descriptive texts

are always displayed during the acquisition in each scenario. The two users have to draw the correct graphical symbols around the texts and connect them using lines, double-lines or arrows. An example of the collected data is shown in 3 (b).



(a) Ground truth of a mind map diagram.



(b) Handwritten diagram without text

Figure 3. Example of a mind map diagram and the corresponding handwritten one without text

A total of 20 people participated in the data acquisition and were partitioned half to half into two groups. Users of group 1 were asked to draw the symbols on a 27" touch screen by fingers. They were given all the freedom to draw the symbol in any way or order they prefer in order to record the way people naturally draw. Since users in this group draw symbols by fingers, some users tend to draw the symbols like 'bracket' and 'double line' and others with two fingers in the same time. In other words, this dataset contains not only multi-stroke but also multi-touch symbols. Users from group 2 made the acquisition on a 80" touch screen by stylus. In this case they have more free space to draw the symbol and may simultaneously draw different symbols very close to each other. Since all the symbols are drawn by stylus, the multi-touch case does not exist in this group. Fig. 1 shows the data acquisition environment from group 2. Since the system only shows the raw ink from the users, the stroke grouping and labeling tasks are achieved manually after data acquisition.

Note that each pair of users were asked to exchange their positions to do the acquisition twice. Finally, a total

of 40 diagrams are obtained in our dataset. The symbols and strokes count are shown in Table I.

Table I
NUMBER OF DIAGRAMS AND NUMBER OF SYMBOLS AND STROKES

Diagrams number	Symbols number	Strokes number
40	2008	2934

B. Diversity of the content

The predefined symbols can be classified into 9 categories shown in Fig. 4. In observing the collected data, users may draw a same symbol with different number of strokes. Examples of rectangle symbol are displayed in Fig. 5. Note that one of the examples shows a very special broken straight stroke case. The User intends to draw a straight stroke, but his finger accidentally lifted up because of the unstable friction on the screen. This situation causes some symbols containing an unusual large number of strokes.

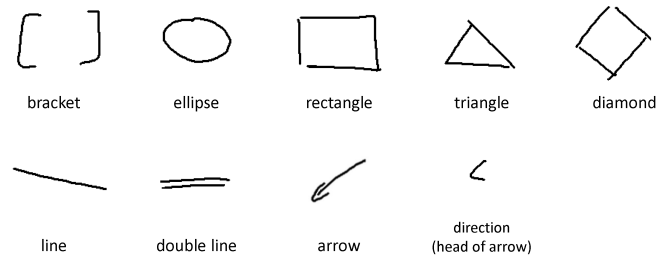


Figure 4. Samples of isolated symbols in diagram.

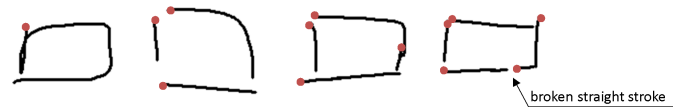


Figure 5. Variability of the rectangle symbol.

As we introduced in the previous section, some 'bracket' and 'double line' symbols may be drawn by multi-touch manner, which means the two strokes of the symbol are drawn simultaneously. Some 'arrow' symbols are written by one stroke, while some others are firstly written as a 'line' and waited a long time delay to have the head added. If the head of an 'arrow' is immediately added just after the straight line stroke, we group the strokes and label the them as 'arrow'. If the head is added after a significant long time delay, during which the body of arrow is supposed to be processed as the 'line', we would make an independent 'direction' class for the head. The heads of double ways arrows are also labeled as 'direction'. Table II shows the distribution of the symbols in the dataset with also the average number of strokes used to draw them.

Table II
SYMBOLS' DISTRIBUTION AND THEIR AVERAGE AND MAXIMUM
STROKE NUMBER IN THE DATASET.

	Symbols	Av. #strokes	Max. #strokes
Bracket	140	2.05	3
Ellipse	278	1.02	2
Rectangle	158	1.89	7
Triangle	88	1.39	5
Diamond	92	1.48	4
Line	446	1.00	2
Double line	140	2.02	3
Arrow	424	1.86	4
Direction	242	1.19	2

C. Format

Each diagram is stored in one file following the XML format InkML¹ from the W3C standard. The raw data only records the strokes from users. InkML offers high flexibility to store and structure the information. Each stroke is preserved in a *trace* element including the x,y and time for each points. Then we use a software to manually group the strokes which, grouped together, form a symbol. Such stroke groups are then labeled with their ground truths using the *trace group* elements.

The device information (Size of the screen) and user information (id, age, gender, writing hand, stylus/finger) are also keep in each file. However, since we don't track the user's input during the acquisition, there is no way to identify the user on the stroke level. The structure of the InkML file is shown in Fig. 6

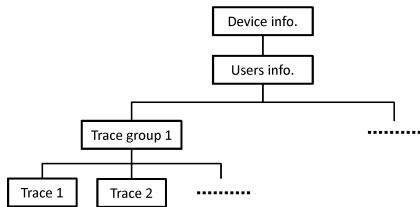


Figure 6. The logical structure of the InkML file.

D. Difficulties for the recognition

The most important feature of this dataset is the mixture of multi-touch and multi-stroke symbols by multiple users in the same time. A key problem is how to correctly group the strokes under this complicated context. Fig. 7

(a) shows an example where a 'rectangle' and an 'arrow' symbol are drawn simultaneously. The diagram in the right describes the temporal activity of their strokes. Obviously, the alternately emerging of the strokes from two users makes it more difficult to group the strokes. Meanwhile, since the symbol is allowed to be written in multi-touch manner, the synchronized strokes (as shown in Fig. 7 (b)) can hardly be determined whether they belong to one user or two. Approximately 65% strokes are simultaneously written in this dataset.

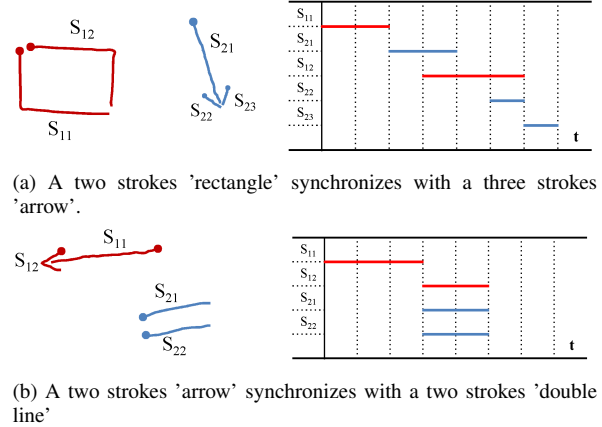


Figure 7. Example of temporal activity of strokes under 2 users condition. S_{ij} indicates the j th stroke from user i .

On the other hand, considering the actual usage of a diagram construction system, the *eager interpretation* is preferred in order to provide the on-the-fly visual feedback to the users. Ideally, a recognition result should be given to the user after a short delay after the completion of each symbol. This is also the reason that we choose to have the independent 'direction' class for the head part of an arrow if the head is added after a long delay. Since there are no clear spatial or temporal boundaries between the symbols in a freely-drawn circumstance, the difficulty is how to find the correct time point to group the strokes from the strokes stream when they can form a meaningful shape. For example, a straight line stroke may be the beginning part of a 'rectangle', 'diamond', 'arrow' or just be the 'line' symbol. The system should analyze the subsequent strokes in the stream within a short delay to decide if it should wait for more strokes to compose another complex symbol or just recognize it as the 'line'.

An example is depicted in Fig. 8. It describes the composition sequence of two symbols drawn simultaneously by two users. At their beginning part, while $t = 300ms$, two straight line strokes are drawn on the screen. Apparently, these two strokes can be interpreted as a 'double line' symbol by the system. When $t = 600ms$, a new stroke emerges on the top right side. The analysis process must be able to decide whether the emerging stroke need to combine with previous

¹<https://www.w3.org/TR/InkML/>

two strokes or is just a independent stroke. In this step the potential interpretation of these three strokes can be a 'line' and a 'direction' symbol. Finally, the two symbols, an 'arrow' and a 'rectangle', are completed at $t = 900ms$. Therefore, the *eager interpretation* may make a different stroke combination and structure interpretation in each step. It is necessary to evaluate if there is enough information to make a decision or wait for forthcoming strokes before interpreting to another symbol.

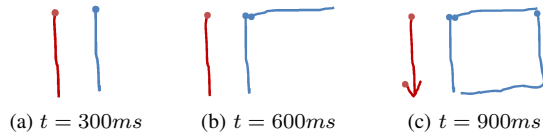


Figure 8. The composition sequence of two symbols drawn by two users.

IV. CONCLUSION

We have presented a new dataset of multi-user handwritten diagram. The dataset contains a large quantity of graphic symbols drawn by 20 users. The most important feature is that each diagram is drawn by the collaboration of two users. The multi-user freely-drawn handwritten recognition is a challenging problem that few recognition systems attempt it. Our dataset opens a new frontier for the diagram recognition research. We present the difficulties to achieve the *eager interpretation* for stroke grouping and symbol recognition in multi-user context. Therefore, new questions are opened: How we can effectively group the strokes into distinct symbols? How we can give a real-time recognition feedback to the user? We aim at using this dataset as benchmark for any further multi-user handwritten recognition system.

ACKNOWLEDGMENT

The authors would like to thank the Excence company from their contribution in the data collection and their financial support of the project.

REFERENCES

- [1] Z. Chen, E. Anquetil, H. Mouchère, and C. Viard-Gaudin, "Recognize multi-touch gestures by graph modeling and matching," in *17th Biennial Conference of the International Graphonomics Society*, Pointe-à-Pitre, France, Jun. 2015.
- [2] D. Rubine, "Specifying gestures by example," in *Proceedings of the 18th Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH '91. New York, NY, USA: ACM, 1991, pp. 329–337.
- [3] J. A. Landay and B. A. Myers, "Sketching interfaces: toward more human interface design," *Computer*, vol. 34, no. 3, pp. 56–64, Mar 2001.
- [4] N. Renau-Ferrer, P. Li, A. Delaye, and E. Anquetil, "The ILGDB database of realistic pen-based gestural commands," in *ICPR2012 - 21st International Conference on Pattern Recognition*, Tsukuba, Japan, 2012, pp. 3741–3744.
- [5] A.-M. Awal, G. Feng, H. Mouchre, and C. Viard-Gaudin, "First experiments on a new online handwritten flowchart database." in *DRR*, ser. SPIE Proceedings, G. Agam and C. Viard-Gaudin, Eds., vol. 7874. SPIE, 2011, pp. 1–10.
- [6] M. Bresler, T. V. Phan, D. Prusa, M. Nakagawa, and V. Hlavác, "Recognition system for on-line sketched diagrams," in *14th International Conference on Frontiers in Handwriting Recognition, ICFHR 2014, Crete, Greece, September 1-4, 2014*, 2014, pp. 563–568.
- [7] C. Carton, A. Lemaitre, and B. Coüasnon, "Fusion of statistical and structural information for flowchart recognition," in *2013 12th International Conference on Document Analysis and Recognition, Washington, DC, USA, August 25-28, 2013*, 2013, pp. 1210–1214.
- [8] S. Macé and E. Anquetil, "Eager interpretation of on-line hand-drawn structured documents: The dali methodology," *Pattern Recogn.*, vol. 42, no. 12, pp. 3202–3214, Dec. 2009.
- [9] M. Nakagawa, K. Machii, N. Kato, and T. Souya, "Lazy recognition as a principle of pen interfaces," in *INTERACT '93 and CHI '93 Conference Companion on Human Factors in Computing Systems*, ser. CHI '93. New York, NY, USA: ACM, 1993, pp. 89–90.
- [10] M. Bresler, D. Prusa, and V. Hlavc, "Modeling flowchart structure recognition as a max-sum problem." in *ICDAR*. IEEE Computer Society, 2013, pp. 1215–1219.
- [11] E. J. Peterson, T. F. Stahovich, E. Doi, and C. Alvarado, "Grouping strokes into shapes in hand-drawn diagrams." in *AAAI*, M. Fox and D. Poole, Eds. AAAI Press, 2010.