

A Novel Online Textual/Graphical Domain Separation Approach for Sketch-Based Interfaces

Danilo Avola, Andrea Del Buono, Pierluigi Del Nostro, and Rui Wang

Abstract Multimodal interfaces can be profitably used to manage the increasingly complex applications and services which support human activities in everyday life. In particular, sketch-based interfaces enable users to effortless and powerful communication way to represent concepts and/or commands on different devices. This kind of information can be expressed by users performing two types of object: free-hand drawing (graphical domain) and/or handwriting (textual domain). Usually, current frameworks require that users, somehow, indicate whether they are performing one or the other object. In this way, the frameworks can adopt the suitable recognition process to interpret as expressed by users. Moreover, more complex situations can occur when users perform, on a same schema, both types of object. This paper describes a novel intelligent framework able to automatically distinguish, in online way, freehand drawing from handwriting. The proposed approach works taking into account only the mathematical features belonging to the sketch performed by the user during interaction activity. Moreover, the approach can be used on schemata made up by heterogeneous objects which can also be overlapped.

Danilo Avola

University of Rome “La Sapienza”, Department of Computer Science, Via Salaria 113, 00189 Rome, Italy, e-mail: avola@di.uniroma1.it

Andrea Del Buono

CSKLab National Research Center, Department of Advanced Research, Via Savoia 84, 00198 Rome, Italy, e-mail: andrea.delbuono@csklab.it

Pierluigi Del Nostro

University of Rome “Rome 3”, Department of Computer Science, Via della Vasca Navale 79, 00146 Rome, Italy, e-mail: pdn@dia.uniroma3.it

Rui Wang

“Saarland” University, Department of Computational Linguistics, Saarbrücken 66041, Saarbrücken Germany, e-mail: wang@coli.uni-sb.de

1 Introduction

Multimodal interfaces allow users to interact, in natural way, with any desktop or mobile devices using multiple different modalities (e.g.: sketch, gesture, speech, gaze, etc.). This kind of interaction provides a powerful tool by which users manage the increasingly complex applications and services that surrounding their activities in everyday life. In particular, sketch-based interfaces enable users to express concepts, to provide commands and to represent ideas in an immediate, intuitive and simple way. In order to express this kind of information users can perform two types of object: freehand drawing (graphical domain) and/or handwriting (textual domain).

The approaches and the algorithms able to detect and to recognize the mentioned objects are deeply different. Moreover, the pre-processing steps that have to be carried out on the sketch before it is given to one or the other recognition engine depends only on the type of object. For these reasons, current frameworks require that users, somehow, indicate whether they are performing freehand drawing or handwriting activity. Besides, it has to be taken into account that complex concepts, commands or ideas could be expressed, by users, using both types of object. In this case, the first step that a framework has to accomplish is to distinguish, within the same schema, what are the objects related to the freehand drawing and what those related to the handwriting. The situation can further be complicated considering that a sketch environment includes several *user actions*, such as: deletes, restyling (of whole object or part of it), and so on. Moreover, it has to be considered that different homogeneous and/or heterogeneous objects can have particular *spatial relationships* with one another, such as: inclusion, overlapping, closeness, and so on.

This paper describes a novel intelligent framework able to automatically distinguish, in online way, freehand drawing from handwriting. The framework, working only on the mathematical features belonging to the sketch performed by the user, is able to overcome all the just mentioned critical duties tied to both *user actions* and *spatial relationships* related to homogeneous and/or heterogeneous objects. Compared to the actual adopted methodologies regarding the separation between freehand drawing and handwriting, our developed approach (and related prototype) has three advantageous main aspects that jointly put our results in the vanguard in the sketch-based interfaces area. The first one regards the possibility to perform both types of object, in online way, on a same schema. The second one concerns the possibility to have both types of object with any possible *spatial relationships* with one another. The last one regards the possibility to perform the separation process without the need to consider specific application domains and/or templates and/or libraries for one or the other type of object. The developed prototype, and the related experimental session, shows the effectiveness of the obtained results.

The paper is structured as follows. Section 2 proposes some remarkable related works which discuss about the separation between freehand drawing and handwriting. Section 3 introduces the novel approach and the related developed framework. Section 4 shows experimental results on a wide application domain made up by

several technical/non-technical freehand drawings and handwriting strings. Finally, Section 5 concludes the paper.

2 Related Works

There are many works about text/graphics separation on different documents, some of these also regarding the online recognition of handwriting (specially on Japanese, Chinese or Arabic language). But no so much works regarding the online separation between objects belonging to freehand drawing and handwriting are given. Besides, no so much experimental results based on an effective framework are shown. However, our work has been inspired from some remarkable approaches described below. A real interesting work, that has driven some our choices regarding mathematical feature extraction process, is shown in [6]. In this work the authors detail a recognition online process able to identify handwritten characters. In order to perform this task the strokes belonging to the characters are treated as curves which have to be suitably classified. A similar work is shown in [7]. The authors model on-line ink traces for a set of several symbols to suitably fit low-degree polynomial series. In order to accomplish this task they use a collection of mathematical writing samples which provides a succinct way to model the stylus movements of current test users. Another remarkable work is shown in [8]. In this work the authors describe a pattern recognition framework able to distinguish global styles of different Chinese people's handwritings. The approach achieves concrete results by combining wavelet transform and generalized gaussian model. Another interesting approach is shown in [3]. In this work the authors propose an offline method to distinguish, within the document, multiple fonts, math notations, and graphics. The strength of proposed work is the possibility to perform document analysis without considering specific structural models, moreover some principles expressed in this work can be suitably adopted in online framework. In [4] the authors propose a method to separate and recognize the touching/overlapping alphanumeric characters in raster-scanned color cartographic maps. The developed approach performs four main segmentation steps, in the first one the map is segmented to extract all text strings including those that are touching other symbols, strokes and characters. In the second one a OCR-based recognition with Artificial Neural Networks (ANN) is applied to define the coordinates, size and orientation of alphanumeric character strings presented in the map. During the third step four straight lines or a number of curves computed as a function of primarily recognized by ANN characters are extrapolated to separate those symbols that are attached. Finally, in the last step, the separated characters input into ANN again to be finally identified. Also in this case several ideas regarding, for example, the overlapped characters, have been deepened. A similar approach, equally useful, is shown in [5]. Also in this work the authors face the separation of overlapping text and graphics in document image analysis. The proposed approach is based on the observation that the constituent strokes of characters are usually short segments in comparison with those of graphics. More than others this approach can be exploited

to support online separation frameworks. Another remarkable work is shown in [9]. In this work the authors present a novel holistic technique for classifying Arabic handwritten text documents. More specifically, the proposed approach exploits both segmentation techniques and feature extraction processes to teach to ANN the classification of styles/fonts need to retrieve Arabic handwritten text documents. From our point of view this work has been interesting “to certificate” some measurement approaches (in particular feature extraction processes) used within of our developed framework.

3 Proposed Approach and Related Framework

This section details the proposed novel approach which works considering only the mathematical features belonging to the sketch performed by the user during interaction activity. In particular, the main algorithmic aspects by which the framework has been developed are shown. In order to explain the section content, next subsection introduces both some basic concepts and a general simplified design of the framework.

3.1 Basic Concepts and Designed Framework

The following definitions and considerations are necessary:

- **Stroke Definition:** A stroke is the set of pixels obtained during the following per action: pen down, pen drawing, pen up.
- **Object Definition:** An object is any sketch, belonging to freehand drawing or handwriting area, made up by one or more than one stroke.

In order to generalize the freehand drawing and handwriting concepts, in this work the following definitions are considered:

- **Graphical Domain Definition:** Belong to this domain all types of freehand drawing. In particular, the drawings can have geometrical and/or non-geometrical characteristics, besides they can represent every object and/or abstraction and/or situation in real life.

Some examples of instances of graphical domains are: dataflow diagram (dfd), entity relationship diagram (erd), electrical schemes, general graphical symbols/icons, geometrical figures, flowcharts diagram (fcd), freehand descriptive drawings, and so on.

- **Textual Domain Definition:** Belong to this domain all types of cursive or block letters strings performed by handwriting. In particular, the strings have to be written according to the following main rules: horizontally (writing from left to

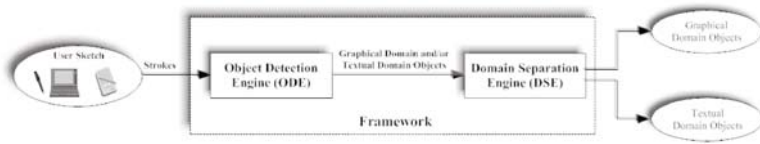


Fig. 1: Design of the Developed Framework.

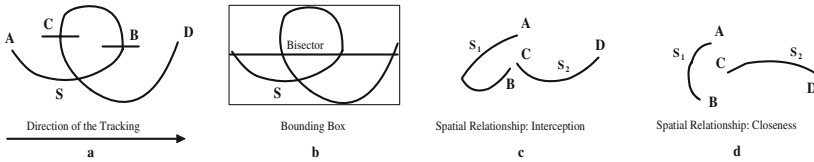


Fig. 2: a) Curvature b) Entropy c) Interception d) Closeness.

right) or vertically (writing from up to bottom). Moreover, the used characters have to belong to the roman alphabet and/or roman numerals.

Some examples of instances of textual domains are: labels on diagrams/schemata, textual description of objects/situations, and so on. In order to describe the main algorithmic aspects by which the framework has been developed in Figure 1 a general design of the developed framework is shown. Our developed framework is made up by two main sequential elaboration engines: Object Detection Engine (ODE) and Domain Separation Engine (DSE). The first one analyzes the set of strokes performed by user, during interaction activity, and identifies how many and what objects the user has sketched. The second one elaborates each single object determining whether it belongs to the graphical or textual domain. The next two subsections detail the just introduced engines.

3.1.1 Object Detection Engine

The aim of the ODE is to identify how many and what objects the user has sketched. In order to accomplish this task on each single stroke and on each set of strokes having particular spatial relationships (*interception* and *closeness*), two main mathematical features are computed: *curvature* and *entropy*. In Figure 2 the just introduced measures are shown. More specifically, as shown in Figure 2-a, the curvature measure represents the sum of the inner angles according to the direction of the tracking of the stroke and/or set of strokes. The stroke S (starting from point A) has two changes of direction (points B and C) and ends to the point D. The inner angles of three sub-strokes ([A, B], [B, C] and [C, D]) are summed to provide the total curvature measure of the mentioned stroke. In Figure 2-b the entropy measure is shown. It represents the messy level of the pixels that make up the stroke,

and/or set of strokes, related to the “bisector” of the bounding box. In particular, the stochastic distribution of the two set of pixels (above and below the bisector) is computed to evaluate the messy level of each set. These two types of measure are always computed on each single stroke, moreover when two or more strokes have particular spatial relationships (*interception* and *closeness*) the mentioned measures are computed on the joined stroke made up by the related set of strokes.

As shown in Figure 2-c two (or more) strokes can be considered intercepting when the start and/or end point of one coincides (taking into account a distance tolerance) with the start and/or end point of the other. While, as shown in Figure 2-d, two (or more) strokes can be considered close when they are near each other (also in this case, taking into account a distance tolerance). Indeed, two or more intercepted strokes are joined together only if they present similar values of curvature and entropy, otherwise they have to be considered only strokes in closeness spatial relationship. Moreover, two or more strokes in closeness spatial relationship are never joined (independently from the curvature and entropy measures).

By the just introduced concepts it is possible to introduce the following definition of homogeneous object able to detect and to distinguish each object sketched by user during interaction activity:

- Homogeneous Object Definition: An object results homogeneous only if it has one of the following main features:
 - High levels of curvature and entropy on each stroke and on the joined stroke;
 - Low levels of curvature and entropy on each stroke and high levels of curvature on joined stroke;
 - High levels of curvature or low levels of entropy on each stroke and high levels of curvature on joined stroke;
 - Medium levels of curvature and entropy on each stroke and high levels of curvature and entropy on joined stroke.

It is possible to obtain some situations in which the found measures do not ensure a correct discrimination of the objects within the schema, in this case it is also considered that users adopt part-by-part strategy, described in [1], to perform any sketch of an object. This strategy states that users tend to end sketching an object before sketching another. Figure 3 shows an example of object detection process. As it is possible to observe ODE, on given schema, detects six different objects (highlighted by the six rectangles shown in the middle window of the screenshot in Figure 3). It is important to highlight that as ODE ends its activity, it provides only how many and what objects the user has sketched, but it does not detect the domain (graphical or textual) of the objects.

Ultimately, it has to be considered that strokes performed in vertical way are “translated” in horizontal during object detection process.

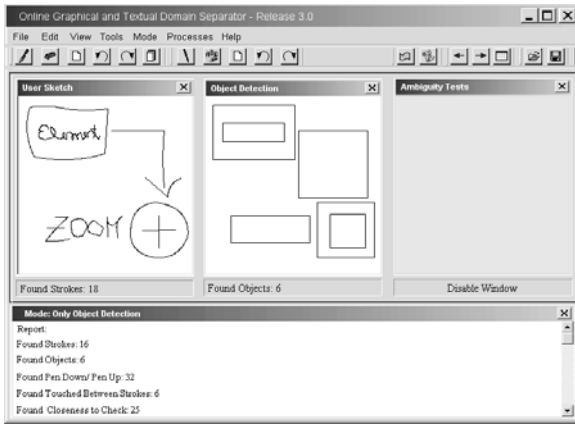


Fig. 3: Example of Six Recognized Heterogeneous Objects on a Schema.

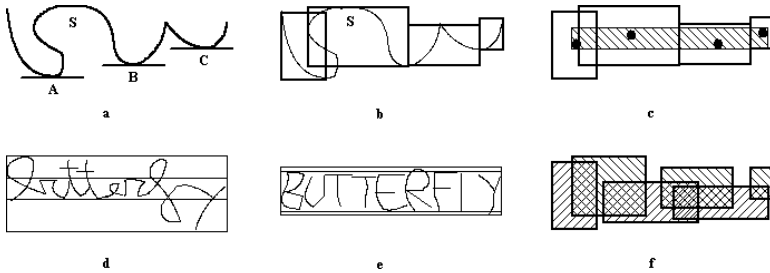


Fig. 4: a) Minimum Points b) Enclosing Rectangles c) Linearity d) and e) Bands Ratio f) Ratio.

3.1.2 Domain Separation Engine

The aim of the DSE is to analyze each object coming from ODE and to detect it as belonging to graphical or textual domain. Indeed, on each object several mathematical features are performed with the purpose to certify the object as belonging to the textual domain. Even if only one measure fails then the object is certified as belonging to graphical domain. In order to explain the measures to analyze each object, it is necessary to introduce the concept of enclosing rectangle built by absolute and relative minimum points of the stroke. Indeed, the mentioned concept is a variation of the approach proposed in [2] which considered both maximum and minimum points. Briefly, as shown in Figure 4-a and 4-b, it is possible to build enclosing rectangles that contain segments of the stroke through the absolute and relative minimum points belonging to the transposition of the two-dimensional space of the stroke on a fixed one-dimensional axes (e.g.: x axes). More specifically, as shown in Figure 4-a, the points A, B and C identify the minimum points (both absolute and

relatives) of the stroke S . By these points, as shown in Figure 4-b, the related enclosing rectangles are built. These rectangles support the definition of the following three main mathematical measures: *enclosing rectangles linearity*, *bands ratio* and *enclosing rectangles ratio*.

The first measure checks the alignment of the enclosing rectangles. As shown in Figure 4-c, on each enclosing rectangle the barycenter point is computed. The enclosing rectangles result aligned if all the barycenter points are horizontally aligned according to specific threshold. The second measure, obtained by density of the pixels within the enclosing rectangles, computes, on a selected object, three density bounds (top, middle and bottom). These bounds identify the behavior of the statistical distribution of the pixels of an object. As shown in Figure 4-d and 4-e these bounds have a high meaning on cursive or block letters strings, more specifically:

- Cursive or Block Letters Strings are characterized by one of the following main features:
 - High density of pixel within the middle bound and equitably distributed low density on both top and bottom bounds;
 - Density of pixels equitably distributed on all three bounds;
 - Density proportionately growing, on the three bounds, in one of the following orders: top-middle-bottom, bottom-middle-top, middle-top-bottom and middle-bottom-top.

In Figure 4-f an example of the last measure is shown. In particular, it regards the ratio between the area obtained from the union of the enclosing rectangles and the area obtained from the intersection of the enclosing rectangles. In particular, according to this measurement, an object is potentially recognized as belonging to the textual domain if the ratio is no more than fifteen percent.

It is important to highlight that an object sketched vertically is, first of all, brought back to the horizontal case according to the tracking verse. As just mentioned, if all three main measures are positively checked on an object, then it can be considered belonging to the textual domain, else it belongs to the graphical domain. Obviously, each measure considers different levels of exceptions depending on: length of strokes, initials and/or finals strokes, strokes made up by small number of pixels (i.e.: points), and so on. Figure 5 shows the domain separation result performed on the example introduced during previous subsection. It might be useful to add another “engine” that merges objects of a same domain having inclusion spatial relationship (in this case the circle with the cruise could be considered a single object). Indeed, this further step depends only on the specific semantic recognition engine used for interpreting the meaning of symbols and/or strings.

4 Experimental Results

Our experimental session has considered sketch activity of two types of schema: simple schema and complex schema. A simple schema was a sketch made up by

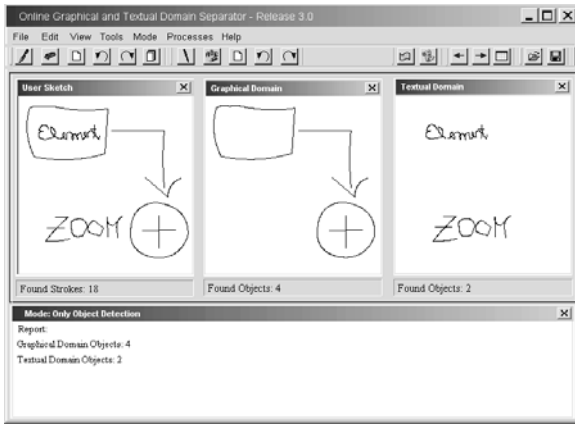


Fig. 5: Four Graphical Domain Objects and Two Textual Domain Objects.

Table 1: Experimental Session.

N° of Simple Schemata (315)	N° of Strokes	Recognized Domain	N° of Objects	N° of Complex Schemata (145)	N° of Strokes	Recognized Domain	N° of Objects
	3	Graphical	2		6	Graphical/Textual	4
	3	Graphical	2		7	Graphical/Textual	4
.....
TREE	9	Textual	1		5	Graphical/Textual	3
gloss	4	Textual	1		5	Graphical/Textual	2
.....
315	2634	Percentage of Recognition 96%	741	145	1513	Percentage of Recognition 89%	912

homogenous objects, which could be either any simple/complex graphical symbol or any simple/complex string. A complex schema was a sketch made up by heterogeneous objects. In this way has been tested each aspect of the two proposed engine (ODE and DSE). In both schemata it was possible to use the desired number of strokes. In Table 1 summarized results of our experimental session are given. As it is possible to observe, test users have sketched 460 schemata subdivided as follows: 315 simple schemata and 145 complex schemata. On simple schemata have been achieved almost optimal results, while on complex schemata the success rate has been lower. This event depends on the error level that occurs during multi-objects separation process in ODE. More specifically, when a user performs several object, on a same schema, can occur that the spatial relationship measures become ambiguous. Moreover, the separation of complex freehand drawings from strings made up by block letters is a hard task.

It is important to highlight following considerations. The schemata have been performed by different users (about 25), this is an important factor because, as well known, each user has its own tracking style that can make the behavior of sketch-based interface oriented algorithms unreliable. The different elements belonging to the tested schemata cover most application contexts, such as: technical diagrams, generic freehand drawings, modeling diagrams, and so on.

5 Conclusion

Sketch-based interfaces enable users to effortless and powerful communication way to represent concepts and/or commands on different devices. This kind of information can be expressed by users performing objects belonging to graphical and/or textual domain. Before using recognition process on each object, it is necessary to detect their related domain. This paper describes a novel intelligent framework able to automatically distinguish, in online way, the related domain of an object. The experimental results, based on developed framework, show the effectiveness of the proposed approach.

References

1. Avola, D., et al.: Ambiguities in Sketch-Based Interfaces. In: The Proceedings of the 40th Hawaii International Conference on System Sciences (HICSS 2007), p. 290b. IEEE Press, Los Alamitos (2007)
2. Avola, D., et al.: Formalizing Recognition of Sketching Styles in Human Centered Systems. In: Apolloni, B., Howlett, R.J., Jain, L. (eds.) KES 2007, Part II. LNCS, vol. 4693, pp. 369–376. Springer, Heidelberg (2007)
3. Changhua, W., Gady, A.: Document Image De-warping for Text/Graphics Recognition. In: Caelli, T.M., Amin, A., Duin, R.P.W., Kamel, M.S., de Ridder, D. (eds.) SPR 2002 and SSPR 2002. LNCS, vol. 2396, pp. 348–357. Springer, Heidelberg (2002)
4. Velázquez, A., Levachkine, S.: Text/Graphics Separation and Recognition in Raster-Scanned Color Cartographic Maps. In: Lladós, J., Kwon, Y.-B. (eds.) GREC 2003. LNCS, vol. 3088, pp. 63–74. Springer, Heidelberg (2004)
5. Cao, R., Tan, C.L.: Text/Graphics Separation in Maps. In: Blostein, D., Kwon, Y.-B. (eds.) GREC 2001. LNCS, vol. 2390, pp. 167–177. Springer, Heidelberg (2002)
6. Golubitsky, O., Watt, S.M.: Online Stroke Modeling for Handwriting Recognition. In: Proceedings of the 2008 Conference of the Center for Advanced Studies on Collaborative Research: Meeting of Minds (CASCON 2008), pp. 72–80. ACM Press, New York (2008)
7. Char, B.W., Watt, S.M.: Representing and Characterizing Handwritten Mathematical Symbols through Succinct Functional Approximation. In: Proceedings of the Ninth International Conference on Document Analysis and Recognition (ICDAR 2007), vol. 02, pp. 1198–1202. IEEE Computer Society, Washington (2007)
8. He, Z., You, X., Tang, Y.Y.: Writer Identification Using Global Wavelet-Based Features. *International Journal on Neurocomputing* 71(10-12), 1832–1841 (2008)
9. Brook, S., Al Aghbari, Z.: Classification of Personal Arabic Handwritten Documents. *Transactions on Information Science and Applications WSEAS* 5(6), 1021–1030 (2008)