Two Decades of GREC Workshop Series. Conclusions of GREC2017

Josep Lladós[✉]

Computer Vision Center and Computer Science Department,
Universitat Autònoma de Barcelona, 08193 Catalonia, Bellaterra, Spain
Josep.Llados@cvc.uab.cat
http://www.cvc.uab.cat

Abstract. This paper summarizes the conclusions drawn in the panel discussion of the Graphics Recognition Workshop (GREC2017) held in Kyoto, Japan in November 2017. As usual, GREC was an exciting workshop with lots of interactions between attendees. Graphics Recognition community is evolving. It is no longer a compact community focused on typical problems as vectorization, text-graphics separation, or symbol recognition. Instead, Graphics Recognition is now a confluence of research problems from different areas with the common interest of interpreting symbolic constructions that follow a context-dependent language.

Keywords: Graphics Recognition · GREC workshop · IAPR-TC10

1 Retrospective: Twenty Years of GREC Workshops

1.1 The Concept of Graphics Recognition

In a traditional view, the field of Document Image Analysis and Recognition has been roughly divided in two major subareas, namely text and graphics recognition. From this point of view where the criterion is the type of information that is extracted from document images, Graphics Recognition can be stated as the subfield of Document Analysis aiming to process documents containing diagrammatic notations. Diagrammatic notations are human communication messages basically consisting of terms such as textual labels, lines and arcs, loops, solid regions, dotted lines, hatched patterns, etc. combined in terms of bi-dimensional rules depending on the domain. Originally, the main categories of graphical documents were engineering drawings, architectural floor plans, and maps. Thus, the main purposes were the conversion of raster images after scanning (large) paper documents into CAD and GIS formats.

1.2   The Evolution of GREC Workshops: A Keywords Perspective

The first edition of the Graphics Recognition Workshop, endorsed by the Technical Committee 10 of the International Association of Pattern Recognition (IAPR), was held at Penn State University, USA, in 1995. Table 1 compiles the intensity of the different contributions in the proceedings regarding the main topics.

Table 1. Papers by Topic in GREC Workshops.

	1995	1997	1999	2001	2003	2005	2007	2009	2011	2013	2015	2017
Low-level processing	16%	7%	0%	10%	6%	6%	0%	0%	8%	13%	0%	0%
Vectorization, primitive extraction, text-graphics	16%	13%	10%	16%	6%	19%	17%	17%	5%	4%	16%	15%
Technical drawings & maps	21%	30%	29%	19%	18%	0%	9%	3%	8%	21%	8%	9%
Layout analysis & diagrammatic notations, music	16%	13%	6%	13%	3%	8%	3%	10%	11%	0%	24%	39%
Applications, systems & architectures	0%	13%	10%	13%	12%	6%	0%	3%	3%	13%	8%	6%
Symbol & shape recognition	11%	13%	23%	6%	18%	25%	14%	17%	18%	8%	8%	12%
Retrieval, indexing & spotting	5%	0%	6%	10%	15%	11%	14%	14%	5%	13%	4%	0%
Sketching, handwritten graphics	0%	0%	3%	10%	18%	8%	11%	10%	16%	13%	12%	3%
Performance evaluation	16%	10%	13%	3%	6%	6%	17%	10%	13%	13%	16%	9%
Historical documents	0%	0%	0%	0%	0%	11%	14%	14%	8%	0%	0%	3%
Camera-based graphics	0%	0%	0%	0%	0%	0%	0%	0%	5%	4%	4%	3%

A first glance analysis of this table leads us to draw the following musings. First, the traditionally considered graphics recognition problems (vectorization, text-graphics separation and symbol recognition) are still there. They are not with the same strength than in the first editions of the workshop, but there is still some research addressed to improve the state of the art, in general in a given context (e.g. symbol recognition in a particular application). We observe an increase in the works on systems for specific document types with diagrammatic notation, in particular tables, flow charts, music scores, etc. This is probably driven by the needs of the market concerning applications for massive reading of certain types of documents. Surprisingly, the traditional document types like engineering drawings, electronic diagrams, maps, etc. seem to decay. These type of documents are nowadays digitally born, therefore the traditional raster-to-vector conversion to import scanned line drawings to CAD and GIS systems is a mature problem from the scientific point of view. Performance evaluation is always present. The community requires standard and open databases and ground truth, and with the increase of the use of machine learning methods, training data is always needed.

Two particular application areas are recovering protagonism: comics and Optical Music Recognition (OMR). We can not consider them genuinely Graphics Recognition problems, and these topics have their own communities. But the links to Graphics Recognition are evident, so they deserve an increasing centrality. It is surprising that sketch-based systems have a low impact in GREC. It is another example of an area of interest that has a research community, but

probably it has stronger ties with the domains of Human-Computer Interfaces, and Computer Graphics than with Document Analysis. It is a challenge for our community in the future, to strengthen the links to this community, and contribute with graphical symbol recognition methods to solve problems of these domains.

1.3   Main Conclusions Drawn in GREC2017

Conclusion 1: In GREC2017 we noticed that. Graphics Recognition is a component in end-to-end interpretation systems (machines as message decoders where graphical languages are an important but not unique component).

The traditional steps (vectorization, text/graphics separation, symbol recognition) are still there but they are losing strength by themselves. However they make sense in a global pipeline. If we analyze them individually, the state of the art is close to consider the problems are solved. The inclusion of traditional topics in a broader context that requires semantic interpretation in a given context (e.g. music scores, diagrams, engineering drawings, maps) is more challenging.

Conclusion 2: Graphics Recognition in more global end-to-end systems. As researchers, there is a need to escape from our comfort zone, where we are designing ad-hoc methods for particular problems. From a semiotic point of view, the field will move from the signifier (recognition of the compounding symbols) to the significant, i.e. the reading and understanding of the sign system in the context where it appears.

There is a need to incorporate more semantics into the process. We are in the artificial intelligence era, where machines understand and act. Graphical objects are understood in terms of a language and a context. There is a need to cope with genericity and heterogeneity, so the systems must learn and adapt themselves to different contexts, not to be designed for ad-hoc for each use case. Graphics Recognition has to be seen as a service that should be offered to several interpretation pipelines. On another hand, systems must be scalable and allow large scale interpretation.

Conclusion 3: Graphics Recognition in the Deep Era. As in textual objets (OCR, HTR, NLP) language models have been integrated in deep learning architectures, the integration of bidimensional language models is a challenge for the next years.

As in the other areas, Deep Neural Networks have irrupted in Graphics Recognition. But is it the silver bullet? Do we really need it for everything? When designing a system, we have to take into account the cost of learning (training data). Graphical documents involve 2D visual languages. In textual input decoding, LSTM+CTC models have been successfully incorporated so they allow to keep memory of the context, i.e. the syntactical structure of the sentence. Graphical constructions usually involve bidimensional languages, which difficult the training process. Paradigms like Graph Neural Networks are promising frameworks to take into account.

Conclusion 4: the need of annotated data. We have to take advantage of the effort made by the community and centralize data and protocols (e.g. the Engineering Drawings Challenge). The role of the TC10/TC11 dataset curators is essencial to define the roadmap for data generation.

A big amount of ground truth data is required, not only for performance evaluation, but also for training. In addition to classical ways of generating data (crowdsourcing) there are new challenging directions to consider: data augmentation, synthetic generation.

2   Current Trends and Challenges

Graphics Recognition is currently present in many problems and applications that involve the interpretation of graphical languages. In addition to the traditional topics that we use to see at GREC workshops, there are interesting problems that are becoming attractive. In this section, we briefly overview these problems and challenges, according to the discussions held during GREC2017.

Graphics-rich document understanding, especially in large-scale scenarios, is a market need. Organizations have digital mail room workflows, where heterogeneous documents, both paper-based and digitally born, have to be processed. The understanding of the contents are required by business intelligence systems. In addition to traditional graphical documents such as engineering drawings, graphical components like logos, stamps, or even tables provide rich information. Components addressed to recognize graphical parts are integrated in ERP and data analytics software.

Flowchart and diagram recognition is a particular type of graphical language that is intensively addressed. Big companies are developing parsing tools for these specific structures. The interpretation of diagrams is useful in different types of applications, as a matter of example, diagrams are efficient communication instruments in scientific papers, in chemical industry, or in patents. In patent interpretation, flowchart interpretation is a useful mechanism to validate or search purposes. A well known challenge for flowchart interpretation in patent documents has been organized since 2009 [3].

The advent of pen or touch-screen based interfaces has increased the interest for **sketch recognition**. Not only for on-line handwriting, which has been a research topic since decades ago, but also for graphical inputs that are the communication language in many emerging applications. The use of sketches in multimodal processing tools has become popular. Sketch-based image retrieval [8] is a growing challenge among the scientific community of computer vision and pattern recognition. Ellis et al. [4] proposed a model that learns to convert simple hand drawings into graphics programs written in a subset of LaTeX.

Doodling in touch screens in smartphones has open a myriad of applications and services. The use of doodles as a simple way to communicate ideas can be used in retrieval, design, education, security, etc. Graphical passwords for user authentificaton is a clear use case that offers flexibility, simplicity and security [6]. Doodling experiences have been proposed online by big companies [1], [2].

These platforms, offered as toy apps, allow to collect many samples from different uses and construct a big ground truth for the community.

Logo Recognition as a particular case of symbol recognition has been one of the central topics of Graphics Recognition. We can observe that beyond the typical application of logo recognition for document classification, there are new applications related to new business services. Brand analysis through social networks is an important issue in marketing departments of companies. An efficient mechanism to track the popularity of the products of a brans is to search for the corresponding logos in the different medias that users publish in social networks. In addition, companies are concerned in forgeries of their brand icons. Scientifically, this is an interesting challenge involving logo detection and classification on the wild. The need of logo databases for training is a crucial need, not only to have instances of real logos but to teach machines to find logos in real scenes. An interesting logo database have been synthetically generated using Generative Adversarial Networks (GAN) [7].

Finally, the literature shows other interesting applications of Graphics Recognition. In [5], Graphics Recognition is used in a multimodal Question Answering system in an educational context. Sixth grade textbooks are analyzed, and the illustrations and diagrams are analyzed together with the textual information. A curious graphics recognition application is graffiti recognition for author identification. It is a forensics problem that has been developed as a tool for Police departments.

3   Final Conclusion and Envisioning the Future

Graphical languages are part of the human communication. Together with textual information, graphical symbols construct messages made by humans to be understood by humans, in the context where they appear. Documents as containers of compound signs, are no longer static paper-based sources, but have evolved to multi-media platforms. Document Analysis has evolved to Reading Systems, in the widest sense. Nowadays, robust reading, sketching interfaces, on-line signature verification, etc. are well-known problems addressed by the document analysis community but they are far from being constrained to process scanned paper documents. The community has open the scope shifting from the object (document images) to the function (interpreting symbols made by humans). Graphics Recognition is aligned with this move. Therefore, the community of Graphics Recognition nowadays is no longer a small but compact group of researchers working on vectorization, text-graphics separation, symbol recognition, etc. but is more a confluence of people coming from different areas (document analysis, computer vision, human-computer interaction, optical music recognition, etc.) that share the interest of interpreting visual (usually bidimensional) languages in their respective fields. Thus, we are now more concerned in methodologies and their application to interpret graphical entities in end-to-end systems.

In conclusion, we see the future of Graphics Recognition as part of global reading systems, i.e. end-to-end systems for interpreting human-made visual

messages. These messages are constructed following a language that is valid in a particular context. The support for these messages can range from the traditional document images to other types of media, including digitally born documents. The Graphics Recognition Workshop held every two years as a satellite event of the International Conference on Document Analysis and Recognition (ICDAR) will attract the interest of researchers from different communities having as common interest the development of techniques for parsing graphical sentences. Methods for graphics recognition will be general enough to adapt themselves to different scenarios and learn incrementally. The need for annotated data will increase in the future, as in other domains of Pattern Recognition. Thus, mechanisms for sharing, compiling, annotating or synthetically generate data will be a relevant focus of attention.

References

1. Autodraw. https://www.autodraw.com/
2. Google quick draw. https://quickdraw.withgoogle.com/
3. CLEF-IP: Retrieval in the intellectual property domain (2009). http://www.ifs.tuwien.ac.at/~clef-ip/
4. Ellis, K., Ritchie, D., Solar-Lezama, A., Tenenbaum, J.: Learning to infer graphics programs from hand-drawn images. CoRR abs/1707.09627 (2017). http://arxiv.org/abs/1707.09627
5. Kembhavi, A., Seo, M., Schwenk, D., Choi, J., Farhadi, A., Hajishirzi, H.: Are you smarter than a sixth grader? textbook question answering for multimodal machine comprehension. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5376–5384, July 2017. https://doi.org/10.1109/CVPR.2017.571
6. Martínez-Díaz, M., Fiérrez, J., Galbally Herrero, J.: Graphical password-based user authentication with free-form doodles. IEEE Trans. Hum.-Mach. Syst. **46**(4), 607–614 (2016). https://doi.org/10.1109/THMS.2015.2504101
7. Sage, A., Agustsson, E., Timofte, R., Van Gool, L.: Logo synthesis and manipulation with clustered generative adversarial networks. CoRR abs/1712.04407 (2017). http://arxiv.org/abs/1712.04407
8. Sangkloy, P., Burnell, N., Ham, C., Hays, J.: The sketchy database: learning to retrieve badly drawn bunnies. ACM Trans. Graph. **35**(4), 119 (2016)