Automatic Content Recognition of Teaching Boards in the Tableau Platform

Conference Paper · January 2012

2 authors:

Gabriel Silva
Federal University of Pernambuco

Rafael Dueire Lins
Federal University of Pernambuco

Automatic Content Recognition of Teaching Boards in the Tableau Platform

Gabriel de França Pereira e Silva and Rafael Dueire Lins
Universidade Federal de Pernambuco, Recife, Brazil
{gfps, rdl}@cin.ufpe.br

Abstract

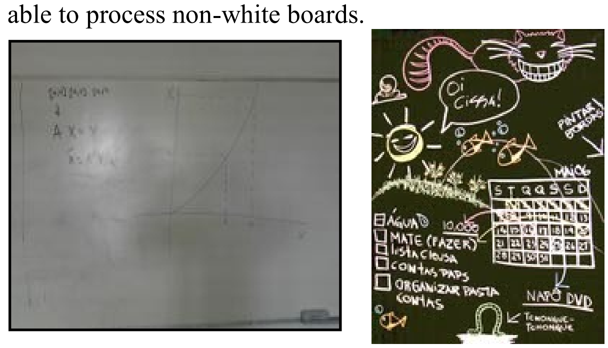
Teaching boards are omnipresent in classrooms throughout the world. Tableau is a software environment for processing images from teaching-boards acquired using portable digital cameras and cell-phones, being the first software environment able to process non-white boards. The aim of the Tableau environment goes far beyond processing and compressing teaching-board images, it also targets at content analysis and recognition. This paper presents the content analysis tool in Tableau that is able to automatically segment and classify textual and pictorial areas.

1. Introduction

Anyone in the teaching area today often sees students in classroom taking photos of the slides being presented or from the teaching board. This phenomenon also happens in conferences and work meetings. It is due not only to the fact that portable digital cameras have become omnipresent, either in their own or embedded in cell phones and tablets, but also because the quality of images acquired has become increasingly better.

Teaching boards are possibly the oldest and most universal didactic tool used throughout the world. Originally, they were made with slices of dark stone to be written on with a piece of chalk. Over the centuries it evolved to whiteboards used with erasable felt pens. Figure 1(a) presents two images of “real world” teaching boards, in which one may observe some of the problems met in such images: perspective distortion, uneven illumination with specular noise, inclusion of background areas and non board elements, etc. Figure 1(b) exemplifies the complexity of a board image in which textual elements are mixed together with drawings, tables, and text in a total absence of pattern and organization.

WhiteboardIt [4, 9, 10, 11] pioneered the research in white board image processing and became a commercial product. Independent work lead to the development of Tableau [19], a simple software tool to process images of teaching boards acquired using portable digital cameras, implemented as an ImageJ plugin, freely available and distributed. Tableau aims to provide a way to generate digital content for courses, respecting particular aspects of the group such as syllabus, class learning speed, teacher experience, regional content, local culture, etc. Although whiteboards have the advantage over chalkboards of generating no chalk dust, that causes allergy to teachers and students, chalkboards are still of widespread use, overall in developing countries. Thus, to reach its original aim to help students and teachers to generate didactic content, Tableau was generalized to process images of any color of board [2]. Until today, this is a unique feature of Tableau, as no other similar tool is able to process non-white boards.



(a) (b)

Figure 1 – Images of a “real world” teaching board.

The Tableau platform consists of three parts. The first is database formation. As soon as images are transferred from the camera to the PC, information is collected to generate a simple database that will organize images for later content formation. Information such as teacher name, course name, discipline, subject, class number, group number, etc. are requested. The second module is for image processing. This module improves the image acquired in a number of ways: background (non-board) removal, image segmentation, skew and perspective correction, image enhancement, etc. The third part of the processing environment deals with outputting the content. Three different ways are under development: printed handouts, webpage generation and Powerpoint™ slide production. Each of these media receives the information of the processed image part of the environment and makes it suitable to its best use. Image "understanding" is needed in order to be able to correctly lay-out the content of the acquired image. This paper targets at such a difficult task and presents the first version of the content analysis tool in Tableau, which is able to discriminate textual and pictorial areas. The text classification is further refined into cursive and block writing, while the pictorial elements are classified into tables, pizza diagram, histogram, and line plotting.

2. Board Graphical Elements

A board image may convey a great variety of graphical elements, such as text, tables, drawings, etc. In general there is no high-level descriptor one could associate with parts of the board image. Trying to understand how such elements are organized in the board image is fundamental to the correct interpretation of its content. The layout analysis has as objective to detect and form the different areas in the image to identify graphical elements as a whole and relation between them. The human brain uses several clues such as contextual information and past experiences about layout recognition together with a sophisticated and complex reasoning mechanism to segment an image classify its elements and "understand" them. The machine, on the other hand, has to infer semantics only from syntax, or in this case from image layout. The idea here is to try to cluster similar pixels forming pictorial elements or text together and "interpret" them by pattern-matching them with "syntactically" similar elements for which semantic value have been attributed to in "artificially intelligent" systems. This is the reason for which automatic layout and structural analysis of arbitrary documents and images is such a challenging research problem. While in some printed documents one may assume some a priori knowledge about layout [3][10][11] such as artifact and graphical settings of textual elements (position of document heading, type of fonts, size, etc.) [3], this is not the case in board images in general.

The system proposed here prioritizes text finding in the board image. Once a cluster of pixels is recognized as text the classification is refined into cursive and block writing. This strategy has shown valuable in making the problem less complex and increasing the correct identification rate.

3. Tableau Board Recognition System

The board element identifier in Tableau encompasses four phases: pre-processing, segmentation, layout analysis and recognition, as sketched in Figure 2.

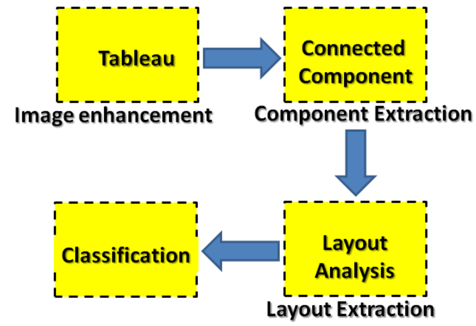


Figure 2 -Block diagram of the element recognition system in Tableau

Image preprocessing is fundamental to the success of the segmentation of the components in the board image. This phase makes use of specific algorithms for noise removal [4] and for the processing of images of teaching boards [2] developed for the Tableau platform. At the end of the preprocessing phase, a binary image is generated by applying the Sauvola-Pietikainen algorithm [5].

4. Layout Analysis and Segmentation

The segmentation step attempts to identify the regions in the board image that encompass graphical or textual elements. The segmentation phase is performed in two steps. First, non-background areas are clustered by using the connected component algorithm. Then, the image blocks formed by clusters are analyzed to be possibly merged by taking into account the average size and distance between graphical and textual elements, defining an envelope-box that should contain the whole element.

The detection of a graphical element should be as coarse as possible. For instance, table recognition must encompass all its elements including the textual information on coordinate names, etc. In general a teaching board has a high density of elements and information, thus the conventional content extraction techniques are not good enough and inefficient. To circumvent such problem a hierarchical information arrangement is adopted here, allowing to link attributes to the extracted content, as shown in Figure 3.

The test set used encompasses 348 "real-world" teaching board images acquired from different courses and classrooms using a portable digital camera manufactured by Sony DSC-T10 of 7.2 MPixels and by the embedded camera in a Samsung Galaxy Mini cell-phone of 5 MPixels. The test images were "hand" labeled and segmented to be used as ground truth in the automatic tests performed. Table 1 presents the number of occurrences of such elements in the dataset.

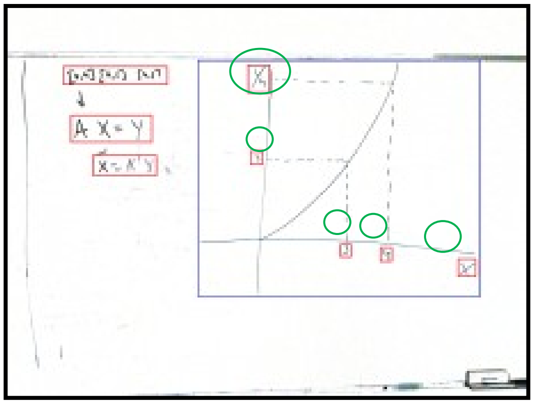


Figure 3 – Hierarchical component connection of the elements of Fig 1(a). In red: textual areas, in blue graphical elements, in green hierarquical association of textual areas to the graphical element.

Type	Total
Tables	136
Cursive Text (words)	2,983
Text in Block format (words)	1,254
Picture	211
Pizza diagrams	321
Histograms	254
Line Plottings	437
Don't know	62

Table 1-Total of classes manually labeled in the test images.

The test images were submitted to the segmentation algorithm described and tested against the "hand-made" segmentation. The accuracy of the segmentation algorithm in finding the "bounding-box" that frames each of the graphical elements in the test images is shown in Table 2

Type	Total	Accuracy
Tables	97	71.32%
Cursive Text (words)	2,132	71.47%
Text in Block format (words)	968	77.19%
Picture	183	86.72%
Pizza diagrams	298	92.83%
Histograms	213	83.85%
Line Plottings	297	67.96%
Don't know	25	40.32%

Table 2 – Accuracy of the automatic segmentation process.

Considering that the test images are "real-world" ones and the teacher did not know *a priori* that the board images would be used for processing, the segmentation rate is good enough, showing that the segmentation tool in Tableau environment is useful.

5. The Element Classifier

In the last processing phase of the Tableau recognition system performs the automatic classification of the graphic and textual elements of the segmented blocks. A cascaded binary classifier was used, in which each of them is a Random Forest [6] classifier. Making the classifier binary allows to reduce the complexity of the classification problem as well as it allows that new branches to be introduced into the classifier in a simple way [12]. The current version of the classifier, as shown in Figure 4, starts by attempting to differentiate between tables and other classes. If a block area is recognized as textual, it goes into another classifier that decides whether the writing is in cursive or in block format. Further classification as a pizza (or pie) diagram, a histogram and a line plotting is made. If the classifier does not "guess" the "nature" of the image block it outputs "Don't_know".

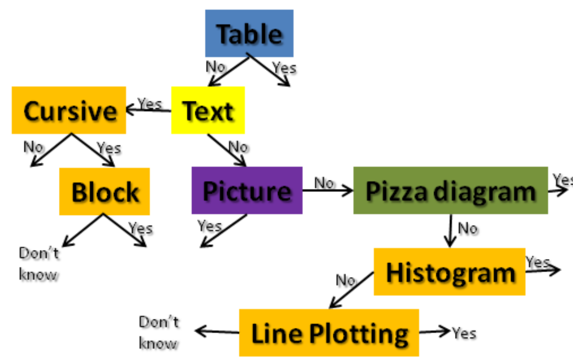


Figure 4 – Tree of binary RandomForest classifiers.

5.1 Feature Extraction

The recent work in image classification [13] points at the model called "bag of visual word" (BoW) [13] based on the quantization of local descriptors [14][15] and SVM (suport vector machines). Such technique is refined with the use of multi-scale spatial pyramids (SPM) [15] together with BoW or with the features of orientation of the histogram of the gradient (HGO) [16]. The current "state-of-the-art" apply multiple descriptors and kernels are combined using learning kernel approaches [17][18]. Such methods did not show suitable to address the classification problem presented here, however. This may be due to the dimension of the data set or to the diversity in the class description universe. Here a Gaussian data distribution is assumed and its performance degrades as their nature distance away from such model. The efficiency in assuming such distributions is reported in references [7][12]. The following features were extracted from each segmented area from the enhanced board image:

- Palette (true-color/grayscale)
- Gamut
- Gamut in Grayscale
- Number of black pixels in binary image.

- (#Black_pixels/Total_#_pixels)*100%
- (Gamut/Palette)*100%
- (true-color/grayscale)
- Number of open and closed loops.
- Size of the segmented block.

The average time for feature extraction was 1.02 s in a Intel® Core 2 Quad Q8300 with 4 GB of RAM.

5.2 Classification Accuracy

The size of the training set used with the different classes is shown in Table 3. The classifier used was Random Forest [6], with cross-validation implemented in Weka [8]. The images in the training set were automatically segmented by the algorithm presented.

Type	Total
Tables	28
Cursive Text (words)	300
Text in Block format (words)	150
Picture	30
Pizza diagrams	60
Histograms	40
Line Plottings	60

Table 3 – Size of the training set.

The performance of the classifier was tested on the manually segmented blocks. No in the training is part of the test set. The classification accuracy is shown in Table 4. The classification time was 7 ms per image.

Type	Total	Accuracy
Tables	87	80.55%
Cursive Text (words)	2,683	91.91 %
Text in Block format (words)	1,104	83.42 %
Picture	131	72.37 %
Pizza diagrams	261	77.01 %
Histograms	214	85.51 %
Line Plottings	377	90.71 %
Don't know	41	66.12%

Table 4 – Classification accuracy for the segmented board elements.

Conclusions

The image segmentation and classifier presented here is an important tool in the Tableau environment, because it will allow it to meet its original purpose and functionality of generating digital content from teaching board images. The automatic segmentation part reached over 70% accuracy, while the classifier reached accuracy over 80% in "real world" teaching board images. It is reasonable to assume that if teachers knew *a priori* that their boards would be processed a better layout would be obtained, increasing the performance of the tool. Another possibility is to customize the tool to a certain teacher, by making his/her "private" training set. Although the Tableau environment aims to ease users' life by inferring the content of teaching board areas, it is recommendable that the user checks the results obtained. This is especially important if the image is to be automatically transcribed to generate a vectorized image.

References

[1] G.F.P e Silva and R.D.Lins. PhotoDoc: A Toolbox for Processing Document Images Acquired Using Portable Digital Cameras CBDAR 2007. p.107 - 115.

[2] D.M.Oliveira, and R.D. Lins. Generalizing Tableau to Any Color of Teaching Boards. ICPR 2010. pp: 2411-2414. IEEE Press.

[3] R.D. Lins, P.H.Espírito-Santo, G.F.P. e Silva. Academus - Generating Digital Libraries of M.Sc. and Ph.D. Thesis. Submitted to ICPR 2012.

[4] G.F.P e Silva, R.D. Lins, J.M.M.Silva, S. Banergee, A.Kuchibhotla, M.Thielo. Enhancing the Filtering-out of the Back-to-Front Interference in Color Documents with a Neural Classifier. ICPR 2010.

[5] J. Sauvola, M. Pietikainen, Adaptive document image binarization, Pattern Recognition 33 (2) (2000) 225–236.

[6] L. Breiman, "Random Forests", Machine Learning, 45(1), pp. 5-32, 2001.

[7] S.J.Simske, Low-resolution photo/drawing classification: metrics, method and archiving optimization, Proceedings IEEE ICIP, pp. 534-537, 2005.

[8] Weka 3: Data Mining Software in Java, website http://www.cs.waikato.ac.nz/ml/weka/

[9] R.D.Lins. A Taxonomy for Noise Detection in Images of Paper Documents - The Physical Noises. ICIAR, Springer Verlag, 2009. v.5627. p.844 - 854.

[10] A. Antonacopoulos, S. Pletschacher, C. Clausner and C. Papadopoulos. Historical Document Layout Analysis Competition, ICDAR 2011, Beijing.

[11] C. Clausner, S. Pletschacher and A. Antonacopoulos. Aletheia - an advanced document layout and text ground-truthing system for production environments, ICDAR 2011.

[12] R.D.Lins, G.F.P e Silva, S.J.Simske, J. Fan, M. Shaw, Mark, M.Thielo. Image Classification to Improve Printing Quality of Mixed-Type. ICDAR 2009. IEEE Press, 2009. p.1106 – 1110.

[13] C.Dance, J. Willamowski, L.Fan, C.Bray, G. Csurka. Visual categorization with bags of keypoints. I. Workshop on Statistical Learning in Computer Vision. 2004.

[14] D.Lowe. Distinctive image features from scale-invariant keypoints. IJCV (2004) 91-110.

[15] S.Lazebnik, C.Schmid, J.Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. CVPR06. (2006).

[16] N.Dalal, B.Triggs. Histograms of oriented gradients for human detection. In: CVPR05. (2005) p. 886-893.

[17] A.Vedaldi, V.Gulshan, M.Varma, A.Zisserman. Multiple kernels for object detection. In: ICCV. (2009)

[18] P.V.Gehler, S.Nowozin. On feature combination for multiclass object classification. In: ICCV. (2009).

[19] D.M.Oliveira and R.D.Lins. Tableau - Processing Teaching-Board Images Acquired with Portable Digital Cameras . CBDAR´2007. IAPR Press, 2007. p.79 - 86