Pattern Recognition 48 (2015) 2129–2140

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

A tree conditional random field model for panel detection in comic images

Luyuan Li [a], Yongtao Wang [a,*], Ching Y. Suen [b], Zhi Tang [a], Dong Liu [a]

[a] Institute of Computer Science and Technology, Peking University, No.128 North Zhongguancun Rd. Beijing 100871, China
[b] Concordia University, 1455 de Maisonneuve Blvd. Montréal, Canada H3G 1M8

ARTICLE INFO

Article history:
Received 14 May 2014
Received in revised form
7 November 2014
Accepted 15 January 2015
Available online 22 January 2015

Keywords:
Comic image document
Panel detection
Probabilistic graphical model
Conditional random field
Context modeling

ABSTRACT

The goal of panel detection is to decompose the comic image into several panels (or frames), which is the fundamental step to produce digital comic books that are suitable for reading on mobile devices. The existing methods are limited in presenting the extracted panels as squares or rectangles and solely use one type of visual patterns, which are not generic in terms of handling comic images with multiple styles or complex layouts. To overcome the shortcomings of the existing approaches, we propose a novel method to detect panels within comic images. The method incorporates three types of visual patterns extracted from the comic image at different levels and a tree conditional random field framework is used to label each visual pattern by modeling its contextual dependencies. The final panel detection results are obtained by the visual pattern labels and a post-processing stage. Notably, the detected panels are presented as convex polygons in order to keep their content integrity. Experimental results demonstrate that the proposed method achieves better performance than the existing ones.

1. Introduction

Comics, defined as "juxtaposed pictorial and other images in deliberate sequence, intended to convey information and/or to produce an aesthetic response in the viewer" [1], have gained increasing popularity since its appearance in the 19th century. During the past few decades, a huge amount of comic publications have been produced all over the world. In this era, comic is inevitably pulled into the road of digitization due to technological advancements. More and more people are willing to read the scanned copies of comic books on cell phones or tablet PCs. In order to cater to the growing needs, the most fundamental and crucial step is to decompose the whole comic page into panels (i.e. panel detection) so that they can be viewed sequentially on the mobile devices. Moreover, the detected panels can be further analyzed to extract features and drawing objects for content based indexing and retrieval of the comic documents.

Comics are popular almost throughout the world. Europe, China and Japan have a lot of comic publications each year. Although the drawing styles among them are disparate, the comic structure is somewhat similar [2]. A comic book consists of several comic pages, which are further split into panels. Each panel consists of a single drawing depicting an important stage of the storyline. It contains drawings such as characters, sceneries and texts (speech balloon, narrative box, etc.). Most panels have discernible enclosing boxes composed of line segments while the others only have drawing objects. At the same time, panels are in many cases separated from each other completely by blank spaces. In other cases, some drawings (such as characters or speech balloons) of one panel are drawn over the adjacent panels. Fig. 1 shows an example of comic page containing the scenarios mentioned above. For some comic pages with rather complicated layouts, on one hand, the panels are still approximate polygons which are very easy for readers to discern; on the other hand, the severe adhesion between the frames makes the task of panel detection rather tantalizing.

In this paper, we present a novel tree conditional random field model for comic panel detection. Within the comic pages, there are several kinds of basic elements that encode structural information of the comic page layouts at different levels. They are named *visual patterns* in our work. Contextual interactions among these visual patterns are of great importance and our novel model has the tremendous ability to model these interactions within a well-constructed framework. As far as we are concerned, we make the first attempt to model the contextual information within the comic pages to address the panel detection task, while the existing methods are some empirical rules designed for certain types of page layouts. We also give an explicit definition on the panels, which preserves the content integrity of the panels better.

* Corresponding author. Tel.: +86 1082529542; fax: +86 1062754532.
E-mail addresses: liluyuan@pku.edu.cn (L. Li), wyt@pku.edu.cn (Y. Wang), suen@encs.concordia.ca (C.Y. Suen), tangzhi@pku.edu.cn (Z. Tang), dong.liu@pku.edu.cn (D. Liu).

http://dx.doi.org/10.1016/j.patcog.2015.01.011
0031-3203/

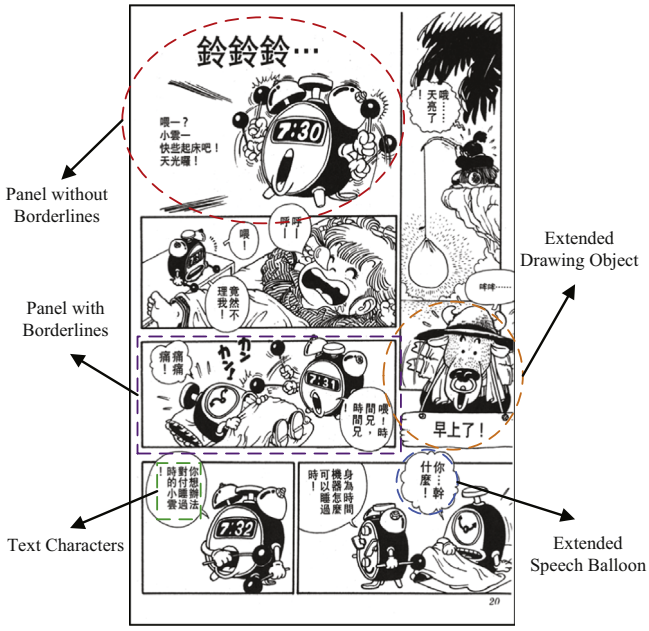

Fig. 1. A typical comic page layout (*Source*: Toriyama, Dr. Slump, Shuisha, vol. 2, p. 20).

The next section of this paper reviews past research works on the panel detection and also image segmentation using structured probabilistic models. Section 3 highlights the importance of context modeling by integrating multiple visual patterns to address the task of panel detection. Our tree conditional random field formulation is presented and described in Section 4. In Section 5, we discuss the implementation details of this panel detection work, including potential and feature design, parameter estimation, model inference and post-processing. We test our method on multiple available datasets in Section 6 and conclusions as well as future works are drawn in Section 7.

2. Related works

2.1. Panel detection methods

Panel detection in comic images (i.e. automatic comic page segmentation) is a relatively new problem which started receiving attention since the publication mentioning the manual panel labeling [3]. Several categories of approaches for panel detection in comic images have been proposed, which can be distinguished according to the type of basic elements extracted from the comic image. To be more specific, the existing method can be classified into three types: division line based [4–7], connected component based [8,2,9] and polygon detection based [10,11].

2.1.1. Division line based methods

The division line based method first proposed by Tanaka et al. [4] iteratively cuts the comic images by horizontal and vertical division lines until the sub-images cannot be divided anymore. In the end of the process, the comic image can be represented by a tree structure. Chan et al. [5] extend the work by focusing on processing color e-comic books consisted of multiple pages. The division line based method is further refined by Yusuke et al. [6], which weights the difference between the contour and other pixels during the division line detection process. Ishii et al. [7] extend Tanaka's work by utilizing the results of frame separation and frame corner detection. The division line based methods are intuitive, but they rely heavily on the results of division lines, which are difficult to acquire when the layout of the comic image becomes complex. Moreover, this method cannot handle the blank margins in the comic image very well.

2.1.2. Connected component based methods

Arai et al. [8] propose another type of methods for comic panel extraction. They define comic images in which no overlap balloons and comic art exist as flat comics and extract these flat comics using modified connected component labeling algorithm. For the overlapped ones, they are separated by line segments. This method cannot handle the situation when multiple panels are agglomerated together by extended objects. Another connected component based method is proposed by Ngo et al. [2]. They first extract connected components from the comic image and then filter them according to their sizes. For the overlapped panels, morphology operations such as dilation and erosion are executed. Such approach is based on a lot of thresholds which are hard to optimize for multiple layouts. Moreover, the morphology operations mentioned in the literature are not very computationally efficient. Recently, Rigaud et al. [9] propose a frame extraction method from comic books. This method processes page per page and begins by acquiring Region Of Interest (ROI) through connected component generation. Then these ROIs are classified as "noise", "text" and "frame" depending on their sizes and topological relations. However, some panels are composed of multiple ROIs and contextual information is not exploited within the framework.

2.1.3. Polygon detection based methods

The polygon detection methods [10,11] see the page segmentation as a polygon detection problem which approximate the panels as quadrilaterals. To obtain polygons, the first step is to detect straight line segments and then combine them to form complete polygons. A post-processing stage is needed to tackle the overlapped ones. Ref. [10] achieves this by designating three empirical rules. Ref. [11] refines the post-processing stage by proposing a line cluster Finite-state Machine to construct the quadrilateral polygons hierarchically. Such approaches make a strong assumption that the frames are all quadrilateral polygons and are enclosed by borderlines composed of straight line segments and thereby would completely fail when the panels do not have enclosing borderlines.

2.1.4. Panel definition

Most of the existing methods do not give an explicit definition on the panels. Therefore, in this section, we formally define the panels in order to clarify the panel detection goals.

In this paper, we present our detection results by defining a panel as a convex polygon linked by a set of convex points that encloses all the contents of the panel. Such definition is based on the following observations:

- Some panels do not have enclosing borderlines.
- Even panels with borderlines do not have a strictly quadrilateral shape.

As shown in Fig. 2, our definition allows a better representation of the panels than that in [11], which is contributed to the ability to include all the contents belonging to a panel. Such character becomes more important in terms of content integrity and better user experience when the readers are reading ordered panels on the mobile devices.

2.2. Image segmentation using structured probabilistic models

Image segmentation using structured probabilistic models such as conditional random field has been widely studied during the