Segmentation-Free Detection of Comic Panels

Martin Stommel[1], Lena I. Merhej[2], and Marion G. Müller[3]

[1] Artificial Intelligence Group, Universität Bremen, Germany
mstommel@tzi.de
[2] Art, Art History and Visual Studies, Duke University, Durham NC, USA,
VisComX, Visual Communication and Expertise, Jacobs University Bremen
l.merhej@jacobs-university.de
[3] Mass Communication, School of Humanities & Social Sciences,
Jacobs University Bremen, Germany
m.mueller@jacobs-university.de

Abstract. The detection of comic panels is a crucial funcionality in assistance systems for iconotextual media analysis. Most systems use recursive cuts based on image projections or background segmentation to find comic panels. Usually this limits the applicability to comics with white background and free space between the panels. In this paper, we introduce a set of new features that allow for a detection of panels by their outline instead of the separating space. Our method is therefore more tolerant against structured backgrounds.

1 Introduction

The understanding of the narrative is one of the most important aspects that artists, art historians and scholars of visual studies pursue in the analysis of iconotextual sequences such as comics and serialized graphics (e.g. poster series or newspaper advertising). The development of dedicated imaging software systems has simplified primarily the technical side of the production process. Tools such as Comic Life (Freeverse Software), Manga Studio Debut (Smith Micro Software Inc.), and Comic Book Creator (Planetwide Games) support the design of graphics and text, online publishing and page layout. Few tools, however, exist to support the analytical side. Audio and video annotation tools such as ELAN[1], ANVIL[2], and KIVI[3] allow for the labelling of the time-line and the creation of multiple views on the data. Unfortunately they do not support the specific structure of a comic where multiple panels with graphical and textual parts (cf. Fig. 1) are arranged in *multiframes* [1] and represented on single or double pages.

A basic functionality of a tool for the analysis of the narration of comics is to automatically extract the panels of the comic pages and display them in reading order. It is the prerequisite for the annotation of the time-line of a comic

[1] Language Archive Organization, http://www.lat-mpi.eu/tools/elan/
[2] Michael Kipp, http://www.anvil-software.de/
[3] http://keyvisuals.jacobs-university.de/kivi.html

L. Bolc et al. (Eds.): ICCVG 2012, LNCS 7594, pp. 633–640, 2012.

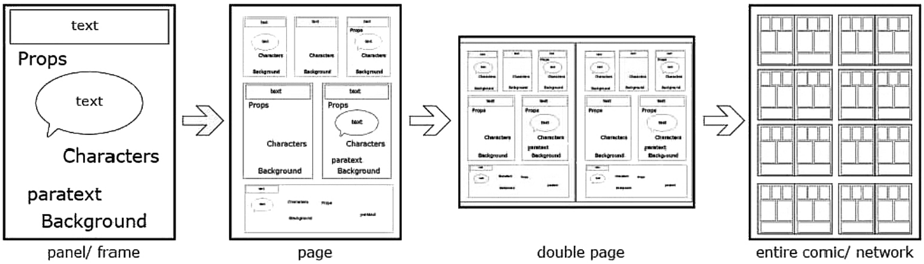


Fig. 1. Structure of a comic book

and therefore the analysis of causal and other dependencies or the rhythm and development of the narrative.

In this paper, we present a method for the automatic detection of comic panels that exploits multiple sources of local and global image information. As a consequence, it is possible to relieve some of the constraints that are usually imposed on the appearance of the panel separations.

2 Related Work

Algorithmic approaches to detect comic panels are related to document structure analysis [2] and image segmentation [3,4]. By assuming that the backround is predominantly white and homogeneous, it can be identified by tresholding, watershed segmentation [5] or region growing [4,6]. The remaining foreground areas are considered as the comic panels. Small connections between panels (caused e.g. by overlapping text balloons) can be broken up to a certain degree by morphological operations [4,6]. For simple layouts, the reading order of the panels can be estimated by first sorting according to the vertical position, then by horizontal position within a row [5].

Since segmentation algorithms are sensitive even to small connections between panels, many systems use top-down approaches [3,4,6]. Pages are recursively split into segments based on detected separating stripes. The lack of visual features of the stripes is compensated by the assumption of a certain length, straightness, and homogeneity. These properties are most frequently measured by axis-parallel [7] or omnidirectional [3] projections of the intensity or the gradient [8]. The ordering of the recursive splits plays an important role, since errors manifest in split, missed, or merged panels. Tanaka et al. [8] therefore introduce a dedicated detector for T-joints of separating stripes. Han et al. [9] improve the noise robustness of the traditional X-Y recursive cut algorithm [10] by reducing the document page to a set of candidate splitting points. The splitting points are detected by a multilayer perceptron. The method is suited for disjoint panels with horizontal and vertical borders. Corner detectors are occasionally used to increase the accuracy of the detected panel positions [11].

3 Proposed Features and Procedural Pathways

As the examples from our data set (Fig. 2) show, the assumption of white background is often not appropriate to find the gaps between the panels. Instead, the background is often coloured. Overlapping text balloons make it difficult to distinguish between panel separations and elements within a panel. We therefore decided to focus on the detection of the panel outlines rather than the detection of homogenous gaps. As a simplification, we assume rectangular panels at the moment. This is of course not always fulfilled. In order to recognise the panel outlines, we developed five procedural pathways that deal with edges, corners, regions, globally dominant structures, and rectangles (cf. Fig. 3).

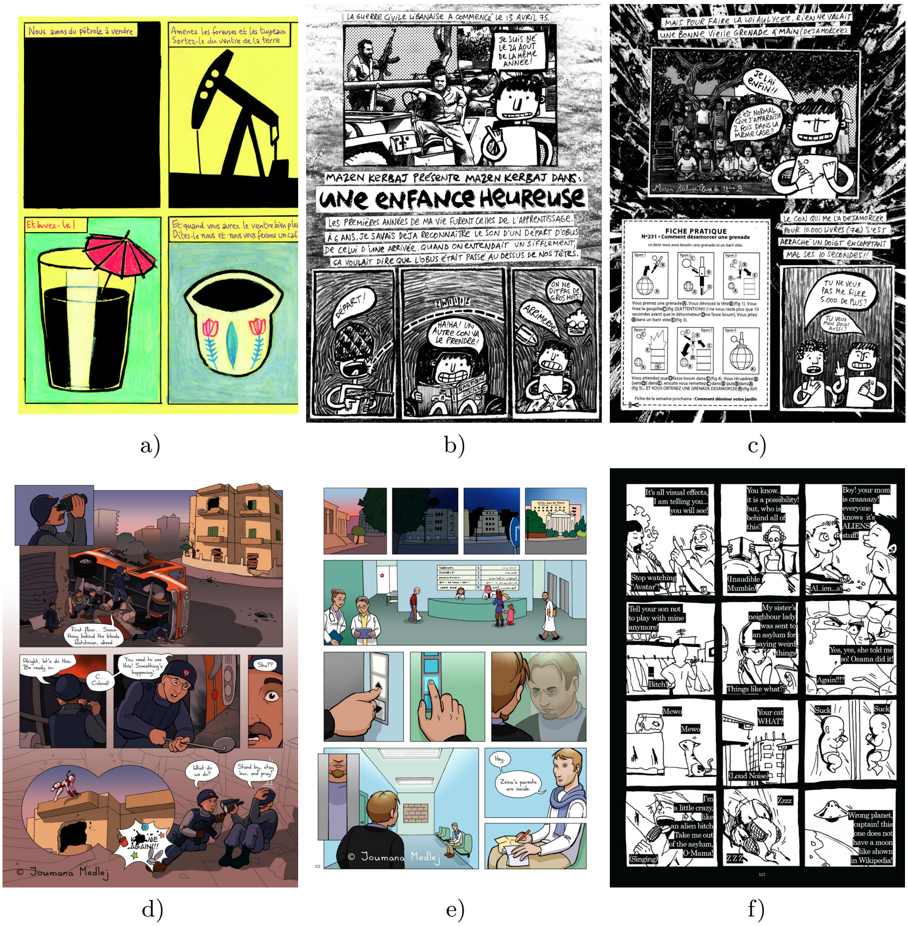


Fig. 2. Some pages from our data set: a) "Cola", b) + c) "Une Enfance Heureuse", d) + e) "Malaak", f) "Mission: Moon"

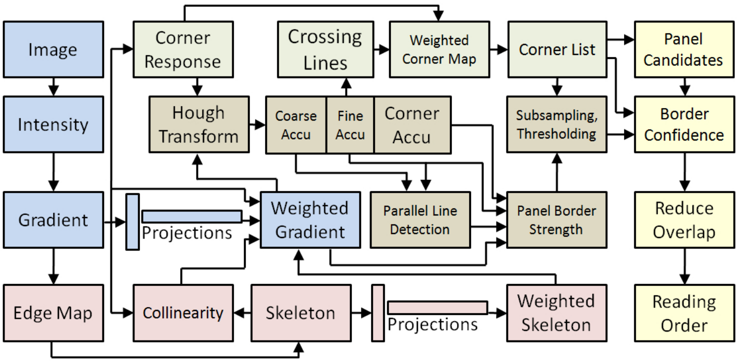


Fig. 3. Procedural pathways of the algorithm

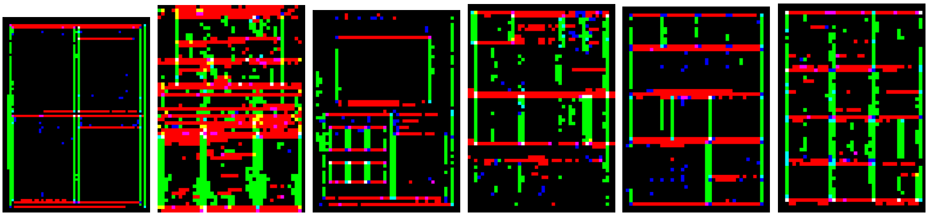
Dominant Vertical and Horizontal Edges. After a grey scale conversion of the input image, the gradient is computed using the Sobel operator (Fig. 3, blue pathway). Then the horizontal and vertical projections of the gradient are computed, i.e. the mean gradient magnitude in each row and column. The aim is to detect rows and columns with long vertical and horizontal edges. To increase the selectivity, only those pixels are considered whose gradient orientation is perpendicular to the direction of the projection. Then for every pixel the gradient magnitude is weighted by the respective average gradient in one line or row, depending on the orientation. The weighted gradient is then scaled by a root function and normalised to the maximum. The two pathways described in the following contribute additional multiplicative weightings.

Stripe-Shaped Gaps between Panels. A skeletonisation is performed to detect homogeneous stripe-shaped gaps between the panels (Fig. 3, red pathway). To this end a coarse edge map is constructed by thresholding the gradient magnitude at half of the maximum. The skeletonisation gives the center pixels of the regions between the resulting edge pixels. For every local region around a skeleton point, the principle orientation of the local skeleton is computed, so we know the orientation of the respective region. We then compute horizontal and vertical projections of the skeleton analog to our procedure in the edge detection. By weighting each skeleton point with the mean number of skeleton points in each row or column, we emphasize horizontal or vertical skeleton lines corresponding to vertical or horizontal stripes. For every edge pixel, we determine the closest point on the skeleton as well as its orientation. The skeletonisation is then used in two ways to reinforce the edge detection results near the panel borders: First, we weight the edge pixel by the weight of the local skeleton. This emphasises edge pixels near horizontal or vertical stripes. Secondly, we weight the edge pixel by the collinearity between the edge direction and the direction of the local skeleton. This suppresses edge pixels which are not the outline of the separating stripe but regular panel contents.

Panel Corners and T-Junctions. The detection of panel corners is solved by a template matching algorithm with additional inputs from a Hough transform [12] (Fig. 3, green pathway). The motivation for template matching was a complete failure of the Foerstner [13] and Susan corner detectors [14]. This might be a result of the coarseness of the comic drawings. Unfortunately, it cannot be solved by downsizing the image because that would destroy strokes and smear gaps. Our template matching algorithm subdivides each local image region into a 3×3-grid and computes the mean vertical and horizontal gradient in each cell. By assigning a positive or negative sign to every cell and orientation and summing up the results of selected cells, we compute the response for the four panel corner directions and four directions of T-junctions. The Hough transform described in the following paragraph provides the accumulated strength of lines running through each corner coordinate. The weights of a possible vertical and a possible horizontal line are read from the accumulator of the Hough transform and multiplied with the response of the corner detector. This prefers corners on the panel borders over corners that lie on short edges or curved contours. A fixed number of the highest ranking local maxima in the corner response are chosen as discrete corner points. In our system the number is set to 80, which is near the maximum number of panel corners on the pages of our data base.

Mutual Reinforcement of Edges and Corners. A Hough transform [12] is used to compute histograms (called accumulators or accu, for short) of hypothetical line parameters (Fig. 3, brown pathway) that are plausible with the local gradient information. We compute accus with fine and coarse bin width. A comparison between the two accus indicates areas with parallel lines, for example in hatchings. Hatchings should be suppressed because they achieve high edge weights but rarely represent panel borders. The fine accu is also used in corner detection as mentioned before. For every line parameter, we also accumulate the corner weights where a line crosses the image. The types of corners accumulated along a line also indicate a panel border: For panel borders near mostly homogeneous stripes, T-junctions and corners to the inside of the panel are more likely. The maximum corner weight that is accumulated for a hypothesised panel border orientation is chosen to weight the gradient magnitude, together with the parallel line suppression. The result is a measure of the panel border strength. The panel border strength under the detected corners is used in an iterative process to threshold the result (see Fig. 4).

Rectangles and Reading Order. In the last part of the procedural pathway (Fig. 3, yellow), the thresholded border strength is converted to a list of rectangles. At first, a list of panel candidates is created by forming a rectangle between all diagonally opposite corners. The confidence of every side of a rectangle is estimated by accumulating the respective values in the thresholded map and searching for suitable corners. Rectangles that cross highly confident borders of other rectangles are discarded, unless they are crossed themselves by other rectangles. Generally unconfident rectangles are discarded. Then every rectangle



a) b) c) d) e) f)

Fig. 4. Thresholded borders (red and blue depending on the orientation) and corners (blue) for the images in Fig. 1

that is intersected by a more confident rectangle is discarded. Intersections with the most confident rectangles are processed first. For the remaining rectangles the reading order is determined. To this end, the geometric relation 'upper left' describing a precedence in reading order is computed for every pair of rectangles. The final reading order is determined by iteratively selecting the first rectangle that is not blocked by other rectangles that have not been selected, yet.

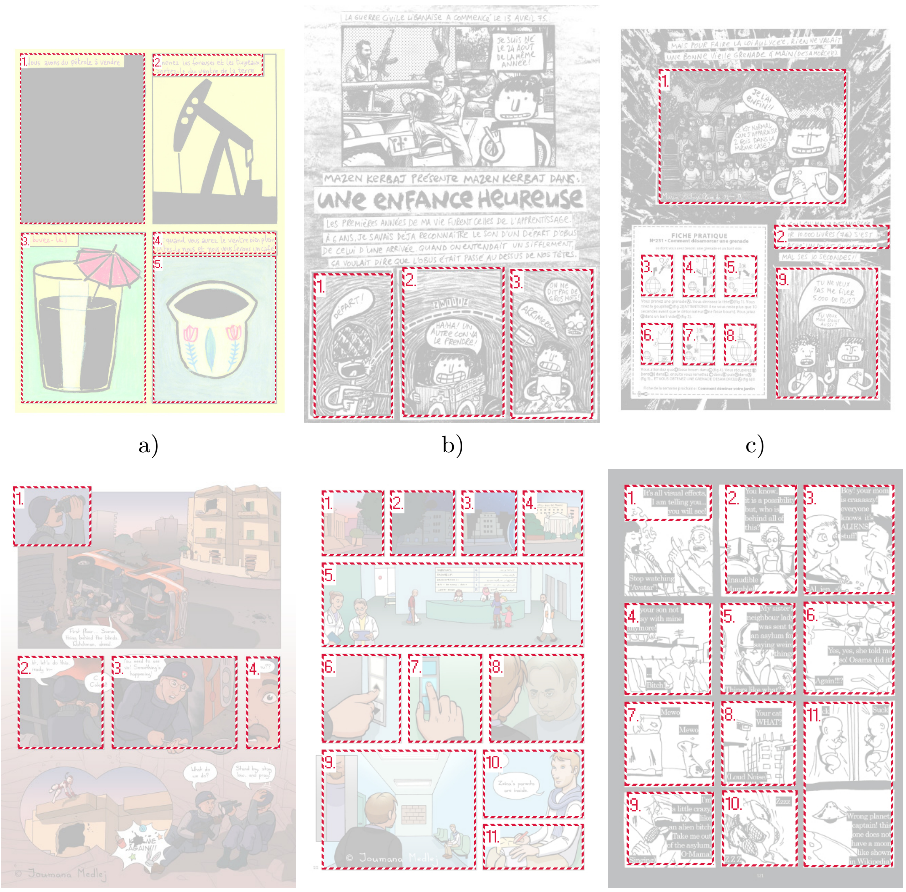
4 Experimental Results

We tested our algorithm on the comics "Cola"[4] and "Une Enfance Heureuse" (2003) by Mazen Kerbaj, "Malaak" (2006) by Joumana Medlej[5], and "Mission: Moon", Ep. 1 by Ahmad Qatato[6] (Samandal Comic Book, issue 8). Figure 2 shows some examples. The combination of edge, corner and skeleton features together with the mutual reinforcement by the Hough transform leads to a robust emphasis of the panel borders with good background suppression. Although the strength of the detection varies over the image plane, the iterative thresholding is usually able to automatically find appropriate binarisation parameters. The corner detector reliably yields the strongest responses for the correct panel corners. Figure 4 shows the output of the feature extraction. False borders are sometimes detected for straight horizontal or vertical lines within the panels. Text balloons extending over the panel borders usually produce gaps in the detections. The same holds for adjoining panels, where the joint border has T-junctions in both directions. This case is difficult to distinguish from other straight lines within a comic.

The analysis of the rectangles is able to resolve most conflicts that are caused by ambiguities and overlaps. As a result, many panels can be localised correctly. Figure 5 shows that most panels outlined by a rectangle have been found correctly. Panels without border do not yield detections (Fig. 5d). A lower recall here is tolerable because it is easier to manually add missing rectangles without having to delete spurious matches beforehand. The reading order is determined

[4] Kerbaj, M.: Cola. In: Le tour du monde en bande dessine, Vol.2, Delcourt (2009).
[5] http://www.malaakonline.com
[6] http://ahmadqatato.com

correctly. However, the filter heuristics are presently too simple to resolve two frequent ambiguities: 1. Small rectangles within a panel are preferred over the surrounding panel outline. This happens for example with rectangular text boxes (Fig. 5a, upper right panel, Fig. 5c, lower left panel). 2. Two adjacent rectangles with a low confidence of the adjoining side are not merged together. The resolution of these conflicts requires better heuristics or information by additional features. The speed of the algorithm is about 5s per image on a desktop computer (Intel Core2 Quad, 2.6GHz). A further speed up is easily possible by employing some of the known standard techniques for code optimisation (e.g. multithreading).



a) b) c) d) e) f)

Fig. 5. Detected panels (dashed rectangles) and estimated reading order (numbers)

5 Conclusion

The detection of comic panels is a difficult problem if the panels are not separated by white background. In this paper we proposed a set of corner based, edge based, region based and global features that allow for the recognition of panels by their outlining rectangle instead of separating white regions. While the heuristics to resolve ambiguities in the assignment of correct rectangles serve basic visualisation needs, our experiments document a high reliability of the feature extraction steps even in difficult image material.

References

1. Groensteen, T.: Systéme de la bande dessinée ("System of Comics", issued in English in 2007 by the University Press of Mississippi). Presses universitaires de France, Paris (1999)
2. Liang, J.: Document Structure Analysis and Performance Evaluation. PhD thesis, University of Washington (1999)
3. Chan, C.H., Leung, H., Komura, T.: Automatic Panel Extraction of Color Comic Images. In: Ip, H.H.-S., Au, O.C., Leung, H., Sun, M.-T., Ma, W.-Y., Hu, S.-M. (eds.) PCM 2007. LNCS, vol. 4810, pp. 775–784. Springer, Heidelberg (2007)
4. Ho, A.K.N., Burie, J.C., Ogier, J.M.: Comics page structure analysis based on automatic panel extraction. In: 9th International Workshop on Graphics Recognition (GREC 2011), Seoul, Korea, September 15-16 (2011)
5. Ponsard, C., Fries, V.: Enhancing the Accessibility for All of Digital Comic Books. Int. J. on Human-Computer Interaction (eMinds) 1(5), 127–144 (2009)
6. Ho, A.K.N., Burie, J.C., Ogier, J.M.: Panel and speech balloon extraction from comics books. In: IAPR International Workshop on Document Analysis Systems (DAS 2012), Gold Coast, Australia, March 27- 29, pp. 424–428. IEEE (2012)
7. Ha, J., Haralick, R., Phillips, I.: Document Page Decomposition by the Bounding-Box Projection Technique. In: IEEE Proceedings of Third Int'l Conf. Document Analysis and Recognition (ICDAR), pp. 1119–1122 (1995)
8. Tanaka, T., Shoji, K., Toyama, F., Miyamichi, J.: Layout Analysis of Tree-Structured Scene Frames in Comic Images. In: International Joint Conference on Artificial Intelligence (IJCAI), pp. 2885–2890. Morgan Kaufmann (2007)
9. Han, E., Kim, K., Yang, H., Jung, K.: Frame Segmentation Used MLP-Based X-Y Recursive for Mobile Cartoon Content. In: Jacko, J.A. (ed.) HCI 2007. LNCS, vol. 4552, pp. 872–881. Springer, Heidelberg (2007)
10. Nagy, G., Seth, S.: Hierarchical representation of optically scanned documents. In: IEEE Comp. Int. Conf. on Pattern Recognition (ICPR), pp. 347–349. IEEE Comp. Soc. (1984)
11. Ishii, D., Watanabe, H.: A Study on Frame Position Detection of Digitized Comics Images. In: Workshop on Picture Coding and Image Processing (PCSJ/IMPS), Nagoya, Japan, December 7, pp. 1–2 (2010)
12. Ballard, D.H.: Generalizing the Hough transform to detect arbitrary shapes. Pattern Recognition 13(2), 111–122 (1981)
13. Förstner, W.: A feature based correspondence algorithm for image matching. ISP Comm. III, Rovaniemi, Int. Arch. of Photogrammetry 26(3/3) (1986)
14. Smith, S.M., Brady, J.M.: SUSAN – A new approach to low level image processing. Technical Report TR95SMS1c, Chertsey, Surrey, UK (1995)