21st International Conference on Pattern Recognition (ICPR 2012)
November 11-15, 2012. Tsukuba, Japan

Robust Stereo Matching for Document Images Using Parameter Selection of Text-Line Extraction

Muhammad Zeshan Afzal[1], Syed Saqib Bukhari[1], Martin Krämer[1],
Faisal Shafait[2], Thomas M. Breuel[1]
[1]Technical University of Kaiserslautern, Germany.
{*afzal, bukhari,kraemer,tmb*}*@iupr.com*
[2]German Research Center for Artificial Intelligence (DFKI), Kaiserslautern, Germany.
faisal.shafait@dfki.de

Abstract

In this paper we present a novel method for automatic text-line parameter selection for stereo image pairs. The parameters are selected such that correspondence between the same content in a stereo pair is maximized. Automatic parameter selection has been carried out by establishing robust text-line correspondence which is also a contribution of the presented work. The proposed method is applied to one text-line extraction algorithm as a proof of concept. The results are compared with the ground truth to show the validity of the method.

1. Introduction

Capturing a high quality model of a book surface is necessary for dewarping algorithms, e.g. [8, 9], which are used to produce flat output from camera captured documents. It is essential to establish robust correspondences between the stereo images in order to obtain a high quality 3D model. However, due to a lot of textureless regions and self similarity of text it is challenging to obtain robust matches.

The importance of text-line correspondence for robust matching has been signified in [1]. The approach uses a naive ordering based method for establishing text-line correspondences which suffers from text-line extraction errors. The work was concerned with showing the improved robustness of 3D matches by using text-line information. It was assumed that the text-lines are almost perfect. In case, if there are some errors in text-line extraction, it was assumed that they are identical for both of the images in stereo pair. This assumption allowed applying a naive approach for text-line correspondences. This assumption does not hold true in most of the cases and suffers from wrong correspondences. This is illustrated in Figure 1. The box at the top of figure shows that one line segment which has been detected in the left image is missed altogether on the right image. A missing component creates ordering problem for the approach presented in [1] which will result in false matches. The second box in the figure shows that some text which belongs to one text-line is considered as a part of another line. It will result either in no match or will produce false matches. The above mentioned problem could be avoided with a technique which works on the content of the page. It is important to mention that if identical text-line extraction errors occur in both of the images then matching procedure will remain unaffected given that one could establish robust text-line correspondences. This paper aims on improving the approach presented in [1] by introducing content based text-line correspondence establishment which is afterwards used for parameter selection of text-line extraction algorithm.

We now introduce a state-of-the-art text-line extraction algorithm proposed by [2] used in this paper. It is based on novel line filter bank which can be tuned to adapt to the structure of the image in combination with isotropic Gaussian filter. The line filter bank has a range of length ($L_{rs} \rightarrow L_{re}$) in pixels and the orientation (θ) as free parameters while Gaussian is characterized by the standard deviation (σ_r). Internally these parameter are computed from the image statistics based on connected components such as average width, height etc.

Intricate line segments are produced when this algorithm is used for stereo images under the same parameter settings due to perspective distortions. It is desired to have a method which performs automatic parameter selection in order to minimize the dissimilarity in the

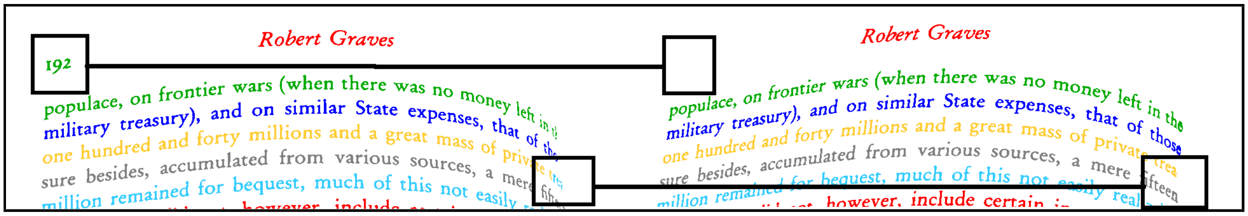


Figure 1: Text-line extraction errors on a small part of the left and right image of the same document captured from different perspectives. The upper boxes connected with lines show that one component is removed. Lower boxes show the incorrect assignment of text-line segments.

segmented content of stereo images.

The presented work has two major contributions. First is the establishment of robust text-line correspondence based on SIFT [5] matching. Each text-line of left image is matched with each text-line of the right image of the stereo pair to obtain a similarity score. The line correspondence problem is solved using the Munkres' [6] version of Hungarian Algorithm [4]. Secondly, It automatically selects optimal set of parameters for text-line extraction algorithm based on the robust correspondence established between the lines corresponding to the stereo pair. To the best of our knowledge, currently there exists no such approach for the selection of parameters for text-line extraction algorithms for stereo images. In order to demonstrate the effectiveness of the proposed method, we compare our results with ground truth quantitatively and the output images are inspected visually.

Next section provides the details of the proposed method which is followed by the experimental results and conclusion.

2. Proposed Method

2.1 Preprocessing

The images are first binarized with Sauvola [7]- a local adaptive thresholding method. A set of values for all the parameters is preselected. The text-line extraction is performed with every parameter contained in the selected set. SIFT features are detected and a bag of features is assigned to every segmented line based on the occupancy of its connected components in the corresponding page as described in [1].

2.2 Feature Matching

The matching between bag of features has been carried out using the similarity measure between the descriptors. Let f be the feature belonging to bag B_{l_i,I_1}, the ith line in the first image and $f^{'}$ belonging to B_{l_j,I_2} be the jth line in the second image the similarity is given by

$$s(B_{l_i,I_1}(f), B_{l_i,I_2}(f^{'})) = ||d_{I_1(f)} - d_{I_2(f')}||_2$$

and matching simply based on this similarity is

$$\mathrm{m}(B_{l_i,I_1}(f), B_{l_i,I_2})) = \arg \min_{f' \in B_{l_i,I_2}} ||d_{I_1(f)} - d_{I_2(f')}||_2$$

where $d_{I_1(f)}$ and $d_{I_2(f')}$ are the descriptors corresponding to feature f and $f^{'}$ from image I_1 and I_2 respectively.

A simple procedure described by [5] is used for rejecting matches. All the matches with distance ratio greater than 0.8 between the first and the second closest neighbor are rejected.

After that matches are cleaned using epipolar constraint

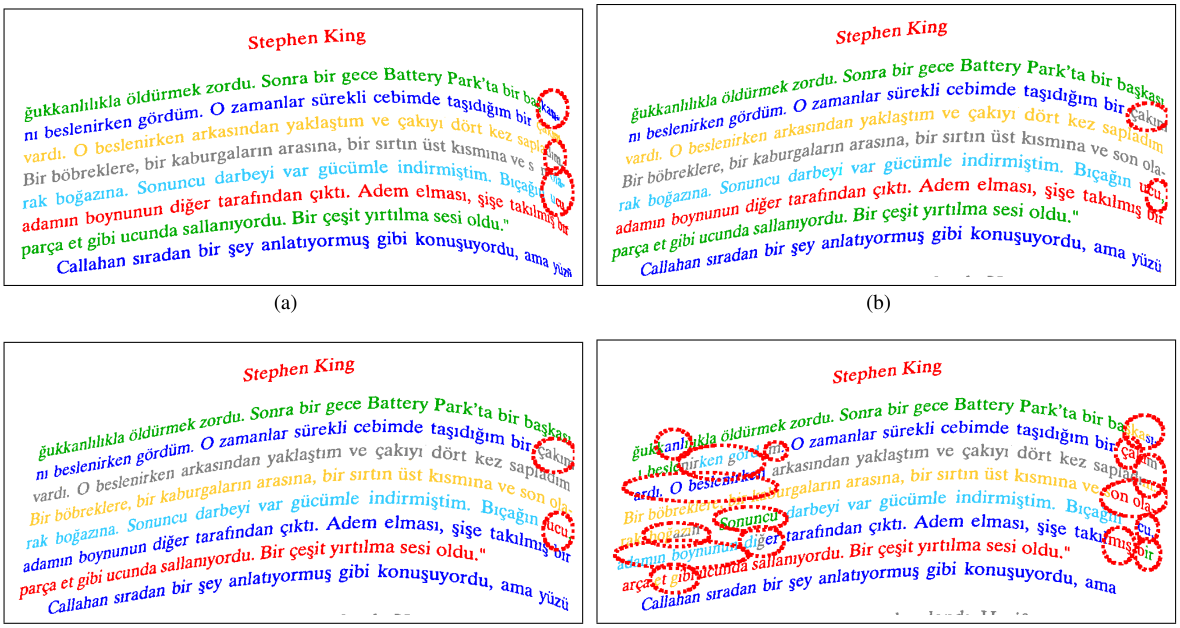
$$q_r^T F q_l = 0$$

where q_l is a point from the left image and q_r its corresponding match in the right image and F is the fundamental matrix. F is estimated from the set of matches by applying RANSAC [3].

2.3 Line Correspondences and Parameter Optimization

In this section we outline a global optimization approach to find the best line correspondences between a stereo image pair. It utilizes the number of detected SIFT features and number of matches for every possible line pairing.

Assume l_L denotes the number of lines in the left image and l_R denotes the number of lines in the right image. Lets further assume that n_l gives the number of detected features in line l of the left image and n_r gives the number of detected features in line r of the right image. The quality of the match of a line pair could be determined by the number of matched SIFT features. Counting only the number of matches could be misleading because a line with very few feature would almost match every big



(a) (b) (c) (d)

Figure 2: Image pairs that achieved the highest and lowest score calculated by the proposed algorithm (a-b) Left and right page (portion) of the pair with highest score. The pair contains few text-line extraction error which are identical in both images (c-d) Left and right page (portion) of the pair with lowest score. It contains a lot of extraction errors which are not identical.

patch due to self similarity of document content. In order to overcome this problem we define a normalizing factor which is the maximum of the amount of information available in both of the lines. This allows us to define a similarity measure f_{lr}, which describes the likelihood, that line l corresponds to line r:

$$f_{lr} = m_{lr} / \max(n_l, n_r)$$

where m_{lr} denotes the number of matches between lines l and r.
Now the similarity matrix

$$S = (f_{lr})_{1 \le l \le l_L,\, 1 \le r \le l_R}$$

can be built, which compares each line of one image with each line of the other image. After building S we normalize it, such that the $0 \le f_{lr} \le 1$ for all l, r. The Hungarian Algorithm solves the assignment problem such that the cost is minimized. Since we are interested in maximizing the score, we invert the values by subtracting each f_{lr} from 1, which allows us to apply the Hungarian algorithm on S for determining the maximum similarity assignment of text-lines.

For parameter selection lets assume we have a set of left images $L = \{L_1, \ldots, L_n\}$ and a set of right images $R = \{R_1, \ldots, R_m\}$,, where different parameters were applied. The idea of the current step is to find the pair $(L_i, R_j)_{1 \le i \le n, 1 \le j \le m)}$ that maximizes the quality of line correspondences as determined by the Hungarian algorithm, i.e.:

$$(i, j) = \arg\max_{i,j} \sum_{l,r} S_{ij}, 1 \le l \le l_{L_i},\ 1 \le r \le l_{R_j}$$

where S_{ij} is the similarity matrix for the image pair (L_i, R_j) as defined in the previous section and l_I denotes the number of lines detected in image I.

3. Experimental Results

The parameters which are used in experimental results are standard deviation (σ_r) and the range lengths of line filter bank i.e. L_{rs} and L_{re}. These parameters are relative quantities based on the statistics of the connected components detected in the image. Some commonly used parameters values are selected for the experiments and are as follows: σ_r, L_{rs} and L_{re} has ranges $[0.25, 0.5]$, $[5, 10]$ and $[10, 15, 20]$ respectively.

We show the results of the parameter selection for the text-line extraction algorithm for the image pairs, with the highest and lowest scores, both visually in Figure 2 and quantitatively in Table 1. The text-lines are colored on the basis of output produced by text-line extraction algorithm. Each color represents a unique line segment. The dotted red circles in the image depict text-line extraction errors. In Figures 2a and 2b a portion of

	Pair with highest score		Pair with lowest score	
Perspective	**left**	**right**	**left**	**right**
Segmented Components (Ground truth)	35	35	35	35
Segmented Components (Image)	35	36	37	53
Under-segmented Components	5	7	8	20
Over-segmented Components	5	6	6	5

Table 1: The comparison of the image pairs with the highest and the lowest score with the ground truth. The pair with highest score is identical to groundtruth in comparison with the one with the lowest score.

the left and the right page of the best selected pair is shown respectively. It can be observed in the upper part of the figure that the best pair has very few extraction errors and the text-lines are almost the same in both of the left and the right images. There are a few errors which are identical and robust establishment of the correspondences is merely affected. Figures 2c and 2d depict a portion of the left and the right page respectively for the image pair with the lowest score. It is clear that this pair is suffering from a lot of extraction errors with high rate of mismatching and hence receives a bad score by the proposed method. Although, the robust line correspondence presented in the paper would compensate for any wrong alignments of text-lines but these areas will be omitted from feature matching step which will lower the overall quantity of number of correct matches.

Table 1 contains the number of ground truth components, number of segmented components, the total over and under segmentation errors. It can be observed that the pair with highest score for the text-line extraction is very close to the ground truth in terms of segmented components and having less segmentation errors. We have inspected visually that the segmentation errors are occurring at the same content and will be aligned correctly by the algorithm. On the other hand, the pair with the lowest score varies a lot in terms of segmented components not only with ground truth but also between left and right images of the pair. Same is the case for the other score provided in the table.

4. Conclusion and Future Work

Text-line extraction algorithms, for example the one used in this paper, have free parameters and need manual adjustment in order to obtain good performance for facilitating feature matching in case of stereo images. This paper has demonstrated that these parameters can automatically be adjusted using robust text-line correspondences. Presented work does not only select the parameters automatically but also provides a way of establishing robust text-line correspondences which is the core component of the proposed method. Future work may include evaluation of text specific features for increasing robustness.

References

[1] M. Afzal, M. Krämer, S. Bukhari, F. Shafait, and T. Breuel. Improvements to Uncalibrated Feature-based Stereo Matching for Document Images by using Text-Line Segmentation . In *Proceedings of the 10th IAPR International Workshop on Document Analysis Systems*, 2012.

[2] S. Bukhari, F. Shafait, and T. Breuel. Text-Line Extraction using a Convolution of Isotropic Gaussian Filter with a Set of Line Filters. In *Proceedings of the 11th International Conference on Document Analysis and Recognition*, pages 579–583, 2011.

[3] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[4] H. W. Kuhn. The hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2:83–97, 1955.

[5] D. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal on Computer Vision*, 60:91–110, 2004.

[6] J. Munkres. Algorithms for the assignment and transportation problems. *Journal of the Society of Industrial and Applied Mathematics*, 5(1):32–38, March 1957.

[7] F. Shafait, D. Keysers, and T. Breuel. Efficient implementation of local adaptive thresholding techniques using integral images. In *Proceedings of the 15th Document Recognition and Retrieval Conference*. SPIE, 2008.

[8] A. Ulges, C. Lampert, and T. Breuel. Document Capture using Stereo Vision. In *Proceedings of the 2004 ACM Symposium on Document Engineering*, pages 198–200, 2004.

[9] A. Yamashita, A. Kawarago, T. Kaneko, and K. Miura. Shape Reconstruction and Image Restoration for Non-Flat Surfaces of Documents with a Stereo Vision System. In *Proceedings of the 17th International Conference on Pattern Recognition*, pages 482–485, 2004.