An Image Consulting Framework for Document Analysis of Internet Graphics

M. KÖPPEN, L. LOHMANN, B. NICKOLAY

Fraunhofer-Institute Production Systems and Design Technology
Sector Process Technology/Department Pattern Recognition
Pascalstraße 8-9, D-10587 Berlin, Germany
e-mail: <mario.koeppen; lutz.lohmann; bertram.nickolay@ipk.fhg.de>

Abstract

A new system approach for image understanding, the image consulting framework, is proposed. It allows for the validation of image properties. Kinds of image properties considered are textual, textural, hierarchically, color and symbolically. Its main application field is information filtering from images used in world wide web documents. The image consulting framework consists of four stages, the color separation stage, the information granulation-verification modules (GVMs), the task stage and the recognition stage. On the base of the framework are the GVMs, which are designed to solve very special tasks. They consists of three parts, a method maintainer, a parameter chooser and a tester. The parameter chooser uses a given set of parameter settings for different runs of the maintained method on the input images of the GPM. The resulting images are tested for the occurrence of the propertyfor which the GVM is designed. All succesfull images are put into a queue. The task stages calls new GVMs due to the filling of the queue and it also assigns input images to the GVMs. All fully treated images are passed to the recognition stage, where the information extraction is performed.

1 Introduction

Image understanding has become one of the most challenging image processing tasks. Past works developed a multitude of algorithms for certain image processing tasks. There are algorithms for e.g. the extraction of edges, ellipses, faces, text regions or document graphics, for the detection of objects, the description of textures or the elimination of backgrounds. See for example in [1], [4], [7], [9]. Also, there is a multitude of families of image processing operations, e.g. morphological operations, filtering, rank operations, chain code evaluation, Hough transformations, texture operators or multiscale operations. The management of these algorithms and operations due to image properties is the task of image understanding.

One restricted application field of image understanding is the extraction of descriptive information from illustrative images, mainly designed and composed by humans due to the principles of aesthetics. The purpose is to combine advertisement and identification. This images can be found in journals, on compact disc covers, front-pages of books and in advertisements.

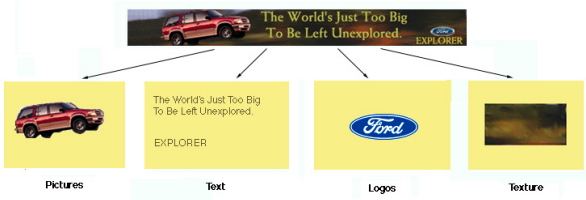


Figure 1: Contents of a browser image

This paper is concerned with a subset of these images, with browser images found in world wide web documents. They serve different purposes, e.g. advertising, referring by hyperlinks to other documents, navigation, simple user interface in separated frames, defaulting ("under construction“, mail box logo, name of the recommended browser), participation in campaigns, corporate design, decoration or server statistics.

On a first glance, many of these images seem to have a very complicated structure. But, one has to consider the functional purpose of these images. In order to fulfill their functional task, these images have to ensure certain aspects. Among these aspects are:

- Detailed information can only be given by written text. Examples for this information are: telephone numbers, product names, dates or names (no logos), navigation goals.
- The image contents is twodimensionally organized. Shadows are the only main instrument for the creation of a pseudo-3d lookout.
- They are not subjected to lightning conditions. Connected regions could be exactly homogenous in color or varying due to exact rules. In general, there is poor reason for noise in browser images. The peculiarities of real-world images can be removed from them before they are published.

- These images are small in size and resolution and have optimized color palettes.

The main information resources of browser images are: text contents, logos, subimage alignment, textures, layout decomposition, coloring, figure-ground separation, geometrical primitives or shadowing (see figure 1). As a consequence, the main concept for the proposed approach is information granulation.

Information granulation is necessary for the treatment of the initially unknown image content. E.g. if one is looking for the text contents of an arbitrary browser image, she can't make any assumptions about the boldness of the character images. Information granulation gives a clue for what to look for. An algorithm uses parameter settings for its calculations. The decomposition of a visual property can be treated by quantization of these parameters („parameter switching"). Afterwards, the result of the algorithms application has to be verified. The interplay of both, information granulation and verification, is on the heart of the proposed approach.

There are many applications for the information obtained from browser images. One is simple to prevent the display of the image in the browser (they may contain unwanted advertisement or be offending). Another application is contents-based image retrieval.

The proposed image consulting framework can be supplied by plug-ins for the world wide web client programs.

In [5] an approach was given for the optical character recognition in browser images. The paper considers the application of document processing techniques for the world wide web. The algorithm given there is a bottom-up approach. It starts from the raw image data and performs a fixed number of steps. A bottom-up approach, from our point of view, does not allow for managing the whole complexity of browser images (see figure 2).

Figure 2: Line histograms over complex background image

In order to improve the performance of an image understanding system for browser images, we overworked it in two ways: treatment of a larger number of visual properties and extensibility by design of additional information granulation-verifying modules (GVM).

The whole framework will be explained in section 2 and an example for its use is given in section 3, followed by an outlook on future work in section 4.

2. The image consulting framework

2.1 Overview of the architecture

The scheme of the image consulting framework is given in figure 3. Data and control flow are separated.

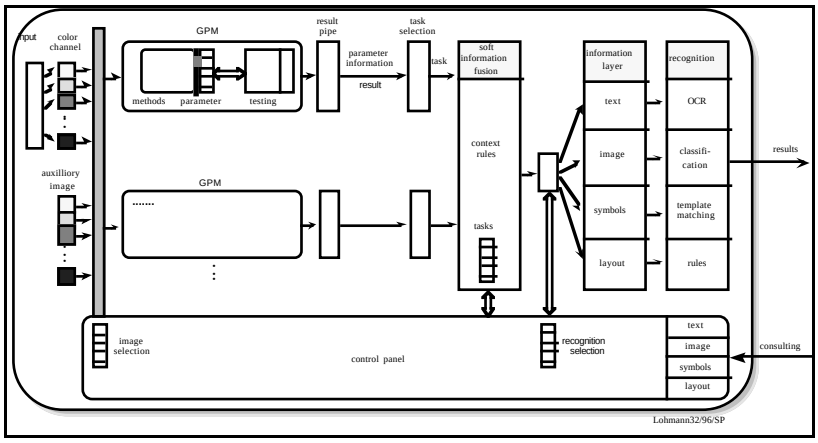


Figure 3: Image consulting framework

The data flow consists of four main stages, the color separation stage, the GVMs, the task stage and the recognition stage.

2.2 Color separation

In general, the processing images are color images. Image processing can not be performed on the color information directly due to the lack of intensity ordering. Therefore, only five color components are used as a result of the color separation stage: the four CMYK components and the saturation component of the HSI decomposition.

2.3 Information granulation-verifying modules

Every GVM consists of three parts, the method maintainer, the parameter chooser and the verification part. The method maintainer performs an algorithm with different parameters settings, which are hold by the parameter chooser. The verifier tests the occurrence of the image property, for which the GVM was designed. The processing flow of a GVM results in a queue of processed images which passed the test. A GVM gets its input from connections to the five intensity images and a set of auxiliary images. The connection can be masked by the processing flow controller. GVMs functionality is re-entrant. Every GVM with at least one input image is activated and starts its processing.

2.4 The task stage

After the GVMs finished processing, the control flow goes over to the task stage. Assigned to every GVM is a task selection unit (TSU). Every TSU has access to a shared resource of tasks. A task is described by a sequence of processing flow controller masks (see 2.3). The queue of images from every GVM is either further processed by another GVMs or it is given to the next stage, the recognition stage.

2.5 The recognition stage

The purpose of the recognition stage is the recognition of given logical levels of information. It is designed as a sequence of layers, e.g. optical character recognition, texture classification, layout decomposition and logo matching layer.

2.6 The processing flow controller

The processing flow starts from the consulting facts given to the control panel. From these facts the needed recognition layer and its expected results are derived by the consultant of the system. The control flow proceeds to the task stage by assessing default starting tasks. GVMs are started by the processing flow controller masks and produce queue images and auxiliary images and selects new masks. If no tasks remains and all queues are emptied, the recognition stage is started. The result of the recognition stage is compared to the consultation goal.

3 Example

As an example, the search item „Netscape“ is extracted from a browser image (see figure 4) by means of the image consulting framework.

Figure 4: Browser image (taken from http://www.netscape.com)

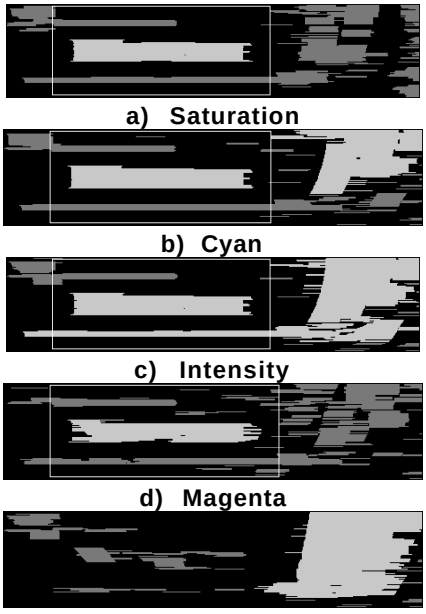
The consultation goal is given as: the character sequence of „NETSCAPE“, all of them in capital letters; imagery of background; and choice of color prefers blue.

3.2 Text location

The GVMs are controlled by bitstrings (in some manner similar to a Turing machine). These bitstrings masks the component and auxiliary images input. Every GVM with at least one input image is activated. The control panel assigns the five intensity images to the GVMs. To identify text contents of browser images, the text location GVM is activated. It maintains an algorithm for text field search, its different parameter settings, and context-based verification methods.

Depending on the result of this GVM, the fusion module will either activate other GVMs (e.g. to check for overlapping text regions or bend text baselines) or pass the extracted text region to the recognition layer, in this case to the optical character recognition layer. This layer performs the further processing of the text region, e.g. binarisation, cleaning, character separation and character recognition by neural network classifier, and classification validation.

The detected text regions in every color channel of the „NETSCAPE“ image are shown in figure 5.



a) Saturation

b) Cyan

c) Intensity

d) Magenta

e) Yellow

Figure 5: Text regions of the browser image „NETSCAPE“

Text regions are detected from their structural properties. An filtering operation (Rhar transformation) enhances the contrast in the image, especially character image borders. The structural property indicating a sequence of character images is the repetition of nearly parallel borders in a connected image region. By the morphological closing operation these borders are connected[8]. By immersion simulation the relevant connected components are extracted. The parameter switching is performed by different parameter settings for the edge operator and structuring elements size. The verification of the extracted regions is performed by evaluating regional features, e.g. features describing the holes inside the connected components.

3.3 Other information sources

To continue the NETSCAPE example, the design of color and logo GVM is sketched.
A color GVM, intended to extract the main color component (if there is any), uses the hue component of the HSI decomposition. Consider figure 6, where the hue image of the „NETSCAPE“ image and its histogram are shown. The histogram has two peaks, one marks red-colored structures from the ship picture, and the other, larger one, marks blue-like structures from the background. The appearance of a peak can be easily verified. The task of the recognition stage is to construct fuzzy terms, in this case: „little red“ and „mostly blue“.

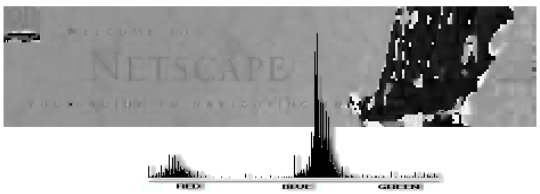


Figure 6: Hue image and its histogram

The logo GVM might use binary morphology for logo template matching. The GVM uses a set of morphological masks which represent indicative substructures for the logo template. The intensity image is binarized by computing the minimal binary representation [6]. After that, the image is eroded with all indicative masks. An indicative mask for the Netscape-Logo can be seen in figure 7.

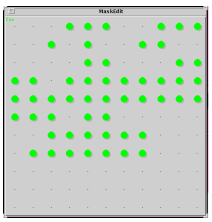


Figure 7: Logo indicative mask

It refers to the lower left corner of the N-image.

4 Summary

A new approach for image understanding, the image consulting framework, is proposed. Its functionality is due to the following design principles: separation of data and control flow; separation of bottom-up and re-entrant algorithms; event-driven task management. Bottom-up algorithms are represented by the recognition layer, re-entrant algorithms by the information granulation-verification modules (GVM).
The functionality of the framework was illustrated by the browser image of the Netscape-homepage. The text location GVM, color GVM and logo GVM have been detailed.
Actually under construction are GVMs for multiscale layout analysis, extraction of geometrical shapes (circles, boxes), extraction of overlapping text regions, extraction of shadows, extraction of straight and curved lines, detection of isolated characters and enhancement of very small characters. A prototype of the image consulting system will be implemented as a plug-in for the Netscape browser and be available in a demo version in the internet.

5 Reference

[1] Anil K. Jain, Sushil Bhattacharjee, Text Segmentation Using Gabor Filters for Automatic Document Processing, Machine Vision and Application (1992) 5:169-184,
[2] M. Köppen, D. Waldöstl, B. Nickolay, A system for Automated Evaluation of Invoices, International Association for Pattern Recognition, Workshop on Document Analysis Systems 1996, p.3 - 21,
[3] D.-J. Lee, S.-W. Lee, A New Methodology for Gray-Scale Character Segmentation and Recognition, Proceedings of the 3th International Conference on Document Analysis and Recognition 1995, p. 524-527,
[4] L. Lohmann, B. Nickolay, System der kanten- und texturorientierten Szenenanalyse am Beispiel der Automatisierung der Umwelttechnik, Tagungsband DAGM 1996, S. 658-665,
[5] D. Lopresti, J. Zhou, Document analysis and the World Wide Web, International Association for Pattern Recognition, Workshop on Document Analysis Systems 1996, p.651 - 671,
[6] M. C. Maccarone, M.Tripiciano, An algorithm to compute medical axis of fuzzy images, Proceedings of The 9th Scandinavian Conference on Image Analysis 1995, p. 525-532,
[7] C. Nastar, B. Moghaddam and A. Pentland, Generalized Image Matching: Statistical Learning of Physically-Based Deformations, Computer Vision- 4th European Conference on Comuter Vision 1996, p. 589-598,
[8] J. Serra, Image Analysis and Mathematical Morphology, Academic Press London 1982,
[9] W. Xinli, F. Albrengsten, Texture analysis using gray level gap lenght matrix, Proceedings of The 9th Scandinavian Conference on Image Analysis 1995, p. 867.