LAMP-TR-062
CAR-TR-959
CS-TR-4212

MDA9049-6C-1250
9802167270
N660010028910/IIS9987944
February 2001

The architecture of TRUEVIZ: A groundTRUth/metadata Editing and VIsualiZing toolkit

Chang Ha Lee and Tapas Kanungo

Language and Media Processing Laboratory
Center for Automation Research
University of Maryland
College Park, MD 20742-3275
{chlee,kanungo}@cfar.umd.edu

Abstract

Tools for visualizing and creating groundtruth and metadata are crucial for document image analysis research. In this paper we describe TrueViz [LK00, KLCB01], which is a tool for visualizing and editing groundtruth/metadata. We first describe the groundtruthing task and the requirements for any interactive groundtruthing tool. Next we describe the system design of TrueViz and discuss how a user can use it to create groundtruth. TrueViz is implemented in the Java programming language and works on various platforms including Windows and Unix. TrueViz reads and stores groundtruth/metadata in XML format, and reads a corresponding image stored in TIFF image file format. Multilingual text editing, display, and search modules based on the Unicode representation for text are also provided. This software is being made available free of charge to researchers.

This research was funded in part by the Department of Defense under Contract MDA0949-6C-1250, Lockheed Martin under Contract 9802167270, the Defense Advanced Research Projects Agency under Contract N660010028910, and the National Science Foundation under Grant IIS9987944.

LAMP-TR-062
CAR-TR-959
CS-TR-4212

MDA9049-6C-1250
9802167270
N660010028910/IIS9987944
February 2001

The architecture of TRUEVIZ: A groundTRUth/metadata Editing and VIsualiZing toolkit

Chang Ha Lee and Tapas Kanungo

The architecture of TRUEVIZ: A groundTRUth/metadata Editing and VIsualiZing toolkit

Chang Ha Lee and Tapas Kanungo

Language and Media Processing Laboratory
Center for Automation Research
University of Maryland
College Park, MD 20742
{chlee,kanungo}@cfar.umd.edu

Abstract

Tools for visualizing and creating groundtruth and metadata are crucial for document image analysis research. In this paper we describe TrueViz [LK00, KLCB01], which is a tool for visualizing and editing groundtruth/metadata. We first describe the groundtruthing task and the requirements for any interactive groundtruthing tool. Next we describe the system design of TrueViz and discuss how a user can use it to create groundtruth. TrueViz is implemented in the Java programming language and works on various platforms including Windows and Unix. TrueViz reads and stores groundtruth/metadata in XML format, and reads a corresponding image stored in TIFF image file format. Multilingual text editing, display, and search modules based on the Unicode representation for text are also provided. This software is being made available free of charge to researchers.

This research was funded in part by the Department of Defense under Contract MDA0949-6C-1250, Lockheed Martin under Contract 9802167270, the Defense Advanced Research Projects Agency under Contract N660010028910, and the National Science Foundation under Grant IIS9987944.

1 Introduction

In the document image analysis (DIA) research area, the term 'groundtruth' refers to various attributes associated with the text on the image — bounding box coordinates of words, lines, characters; font type; character size; direction of text; etc. Groundtruth data is crucial for document image analysis because it is impossible to train and test Optical Character Recognition (OCR) algorithms without it. Since groundtruth is created manually in most cases, tools for annotating and visualizing groundtruth are very important. In fact, at the MLOCR99 international workshop [mlo99] the consensus in the corpus working group was that our community needs i) a protocol for groundtruthing documents, ii) an XML-based groundtruth representation format, iii) a public-domain multilingual/multiplatform visualization and data-entry tool, and iv) a consortium for managing and distributing datasets.

In this paper we address two of the four issues raised by the working group: i) We describe an XML-based groundtruth representation format, and ii) we describe TrueViz, which is a public domain[1] annotation tool that we have developed at the University of Maryland.

This paper is organized as follows. In Section 2 we describe various existing annotation tools used in document image analysis and in related areas such as speech recognition, linguistics, and information retrieval. The desirable features of a document image groundtruthing tool are described in Section 3. In Section 4 we discuss design and implementation issues related to editing, visualization, and search. The XML data format for groundtruth is discussed in Section 5, where we also provide representative samples of XML files. The multilingual data entry, visualization, and search features of TrueViz are quite unique and are discussed in Section 6. Finally, in Section 7 we list the things that we hope the international DIA community will add to the public domain system.

2 Previous Work

There are many annotation and visualization tools in various domains. In this section we describe a few annotation tools commonly used in document image analysis, speech recognition, linguistics, information retrieval, video analysis, geographic systems, and statistics. In Table 1 we provide a comparison of these tools.

2.1 Document Image Visualization Tools

Visualization tools for displaying or editing a document image and groundtruth metadata have been developed for evaluating algorithms, creating document groundtruth, or browsing documents.

Pink Panther [YV98] is an environment for creating segmentation groundtruth files and for page segmentation benchmarking. Page segmentation is the process of decomposing a document page image into structural and logical units, such as images, paragraphs, headlines, tables, etc. The performance of a page segmentation algorithm is evaluated

[1] TrueViz is available at `http://www.cfar.umd.edu/~kanungo/software/software.html`

Table 1: Comparison of Visualization Tools

Name	Platform	Data Format	Domain
PinkPanther	Unix/X Windows System	ASCII	Document Image Groundtruth
Illuminator	Unix/X Windows System	DAFS	Document Image Groundtruth
Oulu Database Browser	Multi-Platform/Java	ASCII	Document Image Groundtruth
TrueViz	Multi-Platform/Java	XML Format	Document Image Groundtruth
Transcriber	Unix/Windows NT	XML Format	Speech Annotation
ATLAS	Unix/Windows NT	XML Format	Linguistic Annotation
Alembic Workbench	Unix system	SGML/PTF Format	Linguistic/Named Entities Annotation
ViPER	Multi-Platform/Java	ASCII	Video Sequence Groundtruth
XGobi	Unix/X Windows System	S Data Format/ASCII	Statistical Data
S-PLUS	Windows 95/98	Customized Data	Statistical Data
CLASP	Unix/Macintosh	Commonly Used Formats	Statistical Data
Mondrian	Multi-Platform/Java	ASCII/Databases	Categorical/ Geographical Data
PolyPaint+	SunOS/Solaris	netCDF	Geographical Data
Spotfire	MS Windows	Database/Spreadsheet/ ASCII	Decision Making by Data Analysis
Slicer Dicer	MS Windows	Binary/ASCII/ Commonly Used Formats	Medical/Scientific Data Defined on Grids

by running the algorithm on a set of document images, and comparing the output for each document to corresponding groundtruth metadata. Pink Panther consists of two parts: Grounds-Keeper and Cluzo. Grounds-Keeper is a tool for creating groundtruth metadata. It visualizes a document image and the corresponding metadata, and also allows users to zone the document image and specify the information for each zone. Groundtruth metadata created by Grounds-Keeper is stored in an ASCII file format. Cluzo is a benchmarking tool for collecting the locations, types and severities of segmentation errors on a page as well as information on segmentation performance. Pink Panther is implemented on the Unix and X Windows platforms and is written in C. While Grounds-Keeper allows the user to enter segmentation groundtruth, entering text groundtruth is not possible.

Illuminator [Fru95] is an editor developed by RAF Technology, Inc. for building document understanding test and training sets, for correction of OCR (Optical Character Recognition) errors, and for reverse-encoding the essential information and attributes of a document. Illuminator visualizes or edits a document image and its entities, which are specific regions of the image and the associated metadata. It is configured to handle text in major European languages and Japanese. Illuminator uses the DAFS (Document Attribute Format Specification) file format [Fru95] to store the document image and metadata. DAFS provides a format for breaking down a document into entities which have hierarchical structure, and for defining entity boundaries and attributes. Illuminator is implemented on the Unix and X Windows platforms and is written in C.

The MediaTeam Oulu Document Database [SK98] is a collection of scanned documents with corresponding groundtruth for the physical and logical structure of the documents. It was developed by the University of Oulu MediaTeam. The document database browser is a visualization tool for exploring the contents of the database. The browser is written in the Java programming language and allows visualization of document images and corresponding metadata simultaneously. The browser can explore the database and select particular documents for visualization. The browser also provides a window to list attributes of the document. Document images which were originally stored in TIFF image format are stored in JPEG image format and metadata is stored in an ASCII file format.

Pink Panther and Illuminator work only on the Unix platform. Because there are many tools that are executable only on the Windows platform, this is a limitation. The Oulu document database browser is written in the Java programming language, and can be run on various platforms. However, the Oulu document database supports JPEG image format only, while TIFF is the most popular image format for document images. Furthermore, the file representation of the groundtruth is non-standard. In fact, all the above tools store document metadata in their own file formats. To provide data compatibility, a standard file format, or a file format to which other file formats can be easily converted, is needed.

A prototype system for visualizing and editing groundtruth is currently being built at the University of Fribourg, Switzerland [HRI00]. This system allows users to edit the hierarchical structure of the document. However, the system does not provide a compatible OCR evaluation package to visualize OCR segmentation results.

2.2 Other Visualization Tools

We surveyed visualization tools in other data domains to find out the best way to provide multi-platform and data compatibility. In this section we summarize features of visualization tools in various domains such as statistical, categorical, geographical, and medical data as well as linguistic data and speech signals.

Transcriber [BGWL00, GBBW00, BGWL98] is a tool for segmenting, labeling and transcribing speech signals. It supports most common audio formats and stores the transcription in XML format. It was developed in the Tcl/Tk and C programming languages, and works on Unix and Windows NT platforms.

ATLAS [BDH+00] is an architecture and tool for linguistic analysis based on a formal model for annotating linguistic artifacts. It uses an XML-based ATLAS Interchange Format (AIF) for storing annotated corpora, and was developed in the C++, Perl, Tcl/Tk and Java programming languages.

Alembic Workbench [DAH+97] is a new set of integrated tools that uses a mixed-initiative approach to bootstrapping the manual tagging process with the goal of reducing the overhead associated with corpus development. The Alembic Workbench is developed using the Tcl/Tk, Perl, C and Lisp programming languages, and works on the Unix platform. Alembic uses the SGML and PTF (Parallel Tag File) formats for source text and annotations.

ViPER (Video Processing Evaluation Resource) [DM00] consists of three main components: ViPER-GT, ViPER-PE, and ViPER-Viz. ViPER-GT contains modules for configuring and producing groundtruth information which describes a video sequence. The ViPER-PE module provides performance evaluation capabilities for comparing computed results with appropriate groundtruth information. ViPER-Viz enables a user to visualize groundtruth, analysis results, performance evaluation results, or an entire video clip. ViPER was developed in the Java programming language, and groundtruth and results are stored in ASCII file format.

XGobi [SCB98, SHB91, SCB92] is an X Window application for interactively exploring statistical data. Its current functionalities include brushing, identification, and editing of connected lines, as well as rotation and the grand tour, with several interactive projection pursuit indices. Several functions can be linked so that actions in one window are promptly reflected in another.

S-PLUS [VR99] is a desktop data analysis tool that provides data analysis and visualization capabilities to identify trends in data. It allows data import and export from spreadsheets such as Excel, as well as from a wide range of relational and other data sources.

The Common Lisp Analytical Statistics Package (CLASP) [AWC+95] is a tool for visualizing and statistically analyzing data. CLASP provides an interactive environment for data manipulation and statistical analysis and a variety of descriptive and hypothesis-testing statistics. It includes many features that facilitate exploratory data analysis.

Mondrian [Uni] is a data-visualization system written in Java. Its main emphasis is on visualization techniques for categorical data and geographical aata. Mondrian provides various plots such as mosaic plots, maps, barcharts, and parallel coordinates, which are fully linked and allow various interrogations.

PolyPaint+ [Nat] is an interactive scientific visualization tool that displays complex structures within three-dimensional data fields. It provides color shaded-surface display, as well as simple volumetric rendering in either index or true color. PolyPaint+ routines first compute the polygon set that describes a desired surface within the 3D data volume, and these polygons are then rendered as continuously shaded surfaces. Objects rendered volumetrically may be viewed along with shaded surfaces. Additional data sets can be overlaid on shaded surfaces by color coding the data according to a specified color map.

Spotfire [AS94] is a decision analysis workspace that uses the connectivity of the Web to provide a workspace in which to access large amounts of complex data from wherever it resides, to visually explore and analyze the data, and to share results.

Slicer Dicer [PIX] provides tools for analysis, interpretation and documentation of complex data defined in three or more dimensions. It helps in exploring the data visually by "slicing and dicing" to create arbitrary orthogonal and oblique slices, rectilinear blocks and cutouts, isosurfaces, and projected volumes. It also provides animation sequences featuring continuous rotation, moving slices, blocks, parametric variation (time animation), oblique slice rotation, and varying transparency.

A more detailed review and taxonomy of visualization tools can be found in an article by Shneiderman [Shn96], and a good general reference for user interfaces is Shneiderman's book [Shn98].

3 Desired GUI Functionalities

Since TrueViz will be used by different researchers for different tasks, we first summarize the functionalities that are desired of such a tool. The simplest task that the tool could be used for is to visualize and input multilingual text. Next, it could be used to mark regions of a scanned document image as text or graphics, and assign labels to regions. A researcher wanting to look at the results obtained by a DIA system might want to search for all the incorrectly recognized characters and then zoom into the image at those locations. A researcher interested in extracting the logical structure of a document might want to label the reading order of the text areas, or the hierarchy of the text regions corresponding to sections and subsections.

After studying the various tasks for which a user might want to use the to-be-designed tool, we formulated the following set of requirements for the graphical user interface:

Entities: Users should be able to visualize and edit zone-, line-, word-, and character-level geometric groundtruth. Furthermore, they should be able to establish their own entity structure. For each entity, they should be able to define attributes (e.g. bounding boxes) and specify their values.

Scale: Users should be able to zoom in and out of the image and overlaid groundtruth so that they can study the image and OCR error results at the page, paragraph, line, word, or character level.

Color: It should be possible to display entities that have different attributes in different colors. For example, image zones could be shown in one color and table or text

zones in another. Thus if a DIA system incorrectly recognizes a table zone as an image zone, the error would be easily identifiable from the color coding.

Logical information: The visualization tool should allow users to visualize and edit the logical reading order of text zones, and also to specify the hierarchy of the text zones. For example, it should be possible to visually specify that a subsection is contained in a section.

Multilingual Visualization: Since DIA systems are being developed for various languages and scripts, users should be able to visualize groundtruth text in these languages and scripts. The use of a standard encoding such as Unicode is highly desirable.

Multilingual Data Entry: While regular English text can be entered by regular keyboards, keyboard mappings that allow other languages and scripts to be entered should also be available.

XML-based Representation: The XML markup language would be ideal for representing page layout groundtruth since it is the current industry standard and various parsers, syntax checkers and editors are publicly available for it.

Converters: Converters to convert standard datasets such as the University of Washington dataset (in DAFS format) into the XML representation would help bootstrap research by providing seed datasets.

Search: Users should be able to search for strings in the groundtruth and find the locations where they appear in the image. The search module should work in any language and users should be able to specify edit distances for approximate searching, which is essential when searching for strings in noisy OCR text.

Evaluation: The tool should have a built-in OCR evaluation module or should be compatible with one, so that users are able to visualize OCR evaluation results easily.

Multiplatform: Since researchers and data entry persons work on various platforms such as UNIX, PC and Mac, the tool should be platform-independent so that users need not spend time learning how to use it on a platform that they are not familiar with.

Public Domain: In order for the community to take full advantage of it, the tool should be freely available.

4 Design and Implementation

4.1 Overview

The TrueViz display is vertically split into two panels (see Figure 1). The left panel is an image panel for displaying a document image and corresponding geometric metadata, and the right panel is a tree view for displaying textual metadata structure.

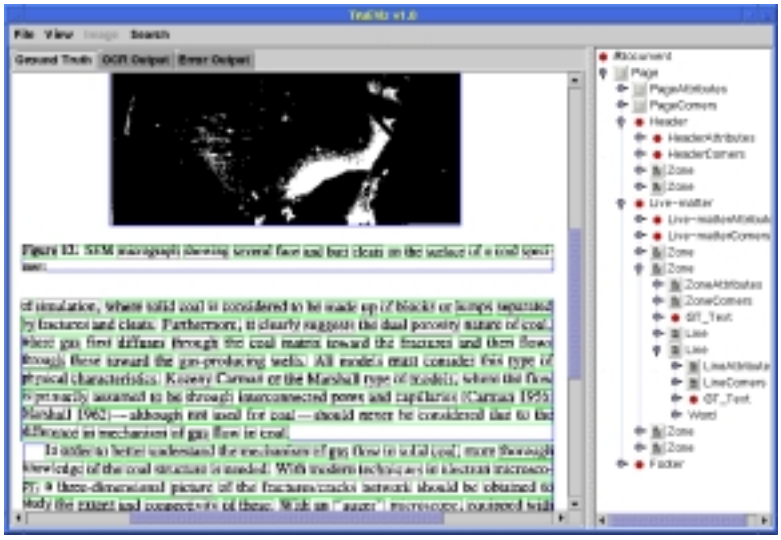

Figure 1: TrueViz consists of an image panel (left) and a tree view (right).

The image panel displays a document image and overlays geometric metadata on the image. Currently, three kinds of geometric metadata can be visualized: Bounding boxes, logical relationships, and an Infopanel. The bounding box of an entity is visualized as a polygon whose color represents the type of the entity. "Logical relationship" refers to logical reading order, and is visualized using an arrow from one entity to the next. The Infopanel is a small window for displaying a few important attributes of the entity. The image and metadata visualization can be scaled to various resolutions.

The tree view displays the XML-based groundtruth metadata in a tree structure of expandable and collapsible nodes. The attribute values can be edited in the tree nodes and the groundtruth text can be edited in the separate multilingual text editor.

4.1.1 Metadata Visualization

Entities can be classified into four categories: Zones, Lines, Words and Characters. Entities are hierarchical in nature, so a Zone is contained within a Page, a Line is contained within a Zone, a Word is contained within a Line, and a Character is contained within a Word. Because of the hierarchical nature of the entities, it is necessary to change views in order to view specific portions of the structure. There are five views: Image Only, Page, Zone, Line, Word and Character. The Image Only view shows only the image without any groundtruth visualization. The Page view shows metadata for all entities, from the highest level to the lowest level. This view is not editable or selectable. The Zone view shows only Zone metadata. A Zone's data can be accessed by clicking on the Zone. This causes the Zone to be active (selected) and highlighted, and the Infopanel to pop up. The Infopanel is a small window for displaying important metadata for the active entity (see Figure 8). The corresponding node in the tree view will also be selected. Similarly, the Line view shows all Line metadata (see Figure 2 (a)), the Word view shows all Word metadata (see Figure 2 (b)), and the Character view shows all Character metadata. As