Release Copy

Lizardtech DjVu Reference

DjVu v3

Document Date: November 2005

From: Lizardtech, a Celartem Company

Status of Standard: Released

1 Introduction

Although the Internet has given us a worldwide infrastmcture on which to build the universal library，much of the world knowledge，history，and literature is still trapped on paper in the basements of the worldfs traditional libraries. Many libraries and content owners are in the process of digitizing their collections. While many such efforts involve the painstaking process of converting paper documents to computer-friendly form, such as SGML based formats, the high cost of such conversions limits their extent. Scanning documents and distributing the resulting images electronically is not only considerably cheaper, but also more faithful to the original document because it preserves its visual aspect

Despite the quickly improving speed of network connections and computers, the number of scanned document images accessible on the Web today is relatively small There are several reasons for this.

The first reason is the relatively high cost of scanning anything else but unbound sheets in black and white. This problem is slowly going away with the appearance of fast and low-cost color scanners with sheet feeders.

The second reason is that long-established image compression standards and file formats have proved inadequate for distributing scanned documents at high resolution, particularly color documents^ Not only are the file sizes and download times impractical, the decoding and rendering times are also prohibitive. A typical magazine page scanned in color at 100 dpi in JPEG would typically occupy 100 KB to 2⑻ KB，but the text would be hardly readable: insufficient for screen viewing and totally unacceptable for printing. The same page at 300 dpi would have sufficient quality for viewing and printing, but the file size would be 300 KB to 1000 KB at best, which is impractical for remote access. Another major problem is that a fully decoded 300 dpi color images of a letter- size page occupies 24 MB of memory and easily causes disk swapping.

The third reason is that digital documents are more than just a collection of individual page images. Pages in a scanned document have a natural serial order. Special provision must be made to ensure that flipping pages be instantaneous and effortless so as to maintain a good user experience. Even more important，most existing document formats force users to download the entire document first before displaying a chosen page. However, users often want to jump to individual pages of the document without waiting for the entire document to download^ Efficient browsing requires efficient random page access, fast sequential page flipping, and quick rendering. This can be achieved with a combination of advanced compression，pre-fetching，pre-decoding，caching，and progressive rendering. DjVu decomposes each page into multiple components (text,

Page 1 of 71