貴重書デジタルアーカイブの構築とその展開

北本朝展　山本毅雄　小野欽司　大西磨希子　村松賢子　池崎友博　佐藤園子　佐藤恵加

デジタルアーカイブとは？

貴重な学術資料・芸術作品などをデジタルデータとすることにより

1. 劣化や破損・消滅を防ぐ
2. 広くアクセス性を高める

さらには、資料の検索性の向上やデジタル修復による再現など、教育・研究に新たな利用法が誕生。

「東洋文庫所蔵」図像史料マルチメディアデータベース

- 国立情報学研究所 (NII) と財団法人東洋文庫との共同研究による、東洋学基本文献（モリソン文庫ほか）のデジタルアーカイブ・プロジェクト。
- 2004 年 4 月に無料一般公開。
- 科学研究費・成果公開促進費などの経費により研究を進めている。

http://dsr.nii.ac.jp/toyobunko/

OCR や翻刻・翻訳テキストの全文検索・連想検索

多言語キーワード検索

地図検索

文章によるコンテンツ解説

メタデータの活用

情報・システム研究機構　国立情報学研究所
〒101-8430 東京都千代田区一ツ橋2-1-2 学術総合センター
TEL (03)4212-2578　E-MAIL kitamoto@nii.ac.jp
http://research.nii.ac.jp/~kitamoto/

貴重書デジタルアーカイブのためのデジタル漂白処理

国立情報学研究所　北本朝展

1．問題の背景

★　貴重書コンテンツをできるだけ便利に活用できるように広く公開したい。

★　それでも、貴重書コンテンツの無断複製はできるだけ防ぎたい。

2．これまでの解決策

★　電子すかしを埋め込むことで、不正利用者を追跡し警告を発する。

→　本気で実施しようとすれば負担も大きい。

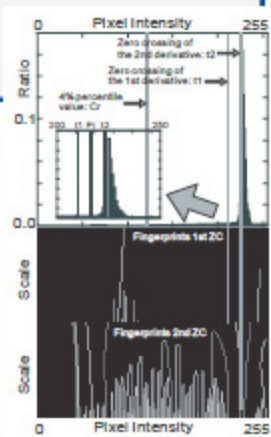
★　複製に価値のない低解像度画像しか公開しない。

→　最も重要なコンテンツであるテキストが読めない。

3．提案する解決策

★　テキスト部分を強調して可読性を高めるとともに、写真など複製価値の高い部分を劣化させるような、画像処理手法を提案する。

→　**デジタル漂白処理 (Digital Bleaching)**



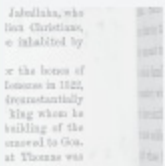
4．デジタル漂白処理

1．紙面領域と文字領域の画素値をヒストグラムから自動推定。

2．紙質の劣化や写真とテキストの混在など、さまざまな変動要因に対しても頑強な手法として、**スケールスペースヒストグラム (scale space histogram) 法**を利用する。

3．東洋文庫アーカイブ 9062 枚の画像の全自動処理を実現。

5．実験の結果



脱色処理　漂白処理

脱色処理　漂白処理

脱色処理　漂白処理

情報・システム研究機構　国立情報学研究所
〒101-8430 東京都千代田区一ツ橋2-1-2 学術総合センター
TEL (03)4212-2578　E-MAIL kitamoto@nii.ac.jp
http://research.nii.ac.jp/~kitamoto/