

# PDF による超高压縮カラーコンテンツ 作成効率化システムの開発

The development of the PDF based high compressed full color contents generation support system

本田 克己<sup>1)</sup> 芦塚 浩一<sup>2)</sup>  
Katsumi HONDA Koichi Ashizuka

- 1) 株式会社ハイパーギア  
(〒349-0123 埼玉県蓮田市本町2丁目19番 渡辺ビル E-mail: hg-honda@po.ijnet.or.jp)  
2) 株式会社ハイパーギア 技術部  
(〒349-0123 埼玉県蓮田市本町2丁目19番 渡辺ビル E-mail: ashi@sb3.so-net.ne.jp)

**ABSTRACT.** Although it is important, converting paper information such as a catalog into the electronic media for the Internet is difficult because of the data size. The HC-PDF technology which is based on the de facto standard of an electronic document PDF, compresses character and picture portion separately, so that reduce data size dramatically, but the productivity is not good because of manual process of the graphics. We developed two Linux based systems, one for graphics workstation which separate character, picture portion and compensate scanned image, and one for ASP server which convert images to HC-PDF, so that at least twice productivity improvement can be gained

## 1. 背景

インターネットを用いた電子商取引の分野においては、ユーザに提供できる情報の量が取引の成否を決める鍵であるが、ウェブ上では要約された比較的容量の小さい情報しか提供できなく、このため、取引の機会を逸したり、ユーザは別途郵送などの手段で紙のカタログを送付してもらうなど不便を強いられるかたがとが現状である。

当社では、これらの問題を電子文書のデファクトスタンダードである PDF を利用して解決する HyperCompact PDF (HC-PDF) を開発した。これは標準の PDF フォーマット上で、画像部分と文字部分を PDF のマルチレイヤーの最適化した画像ファイルとして合成することにより、例えば旅行会社のパンフレットなど、FAX でもつぶれてしまう程細かい文字を含むカラーの印刷物でも、1ページ100-200KBと充分実用になるサイズに高压縮する技術である。これにより高速の回線をもたない、一般のインターネットユーザでも、1ページ20-40秒程度で細かなパンフレットなどをいつでも取出し、安価なインクジェットプリンタでも1分程度で印刷することができる。しかしながら、高压縮を実現するためには、自動処理だけでは変換できないという点があげられる。汚れ、ゆがみ、網点処理の補正や、文字部分と画像部分の分離など手作業による補正、調整が1ページあたり2,3時間かかり、またそのためには高度な画像処理ソフトウェアが使えるデザイナーが必要など、限られた人しか作業できなかった。このように、HC-PDF形式は、小さい容量で詳細な情報を提供できるデータ形式であるが、紙の印刷物をスキャンして得たデータからHC-PDF形式のデータを作成するツールとして、画像と文字を効率的に分離処理する機能を含むソフトウェアの開発が期待されている。

## 2. 目的

本開発では、紙の印刷物をスキャンして得た画像データから、画像と文字を分離処理する機能を含む画像処理システムと、分離した各々のデータからHC-PDF形式のデータを生成するサーバーシステムを開発するものである。本開発では、画像の補正を極力自動化し、また文字部分や画像部分の分離もLinuxで、誰でも使える専用のツールを用意することにより、これらを在宅の主婦などで可能にするサービスビジネスとして展開できるようにして、世の中にある多くのカタログや、パンフレットなどを極めてリーズナブルなコストで電子化できる仕組みを提供することを目的とする。また補正などのツールに加えて、多量の計算能力を必要とするPDF変換部分は、HC-PDF変換サーバをASP(Application Service Provider)として用意することにより、各作業員毎に高価な設備投資をしなくてもすむ仕組みを開発する。これにより、一連の作業を、家庭の主婦や障害者を含むSOHO事業者と協業して実施することにより、新たなビジネスモデルの創出を目指すものである。

## 3. 開発内容

### 3.1 画像処理方式の開発

本開発では、スキャンされた画像データから、極力自動処理により、効率よく、画像と文字の分離を行い、また、スキャナ入力による品質の低下などを補正するために、以下の開発項目に対して、その実装方式を検討して、ソフトウェアとして実現し、その結果を評価するものとする。

#### (1) 文字画像分離方式の開発

カタログやパンフレットを含む印刷物をスキャンして

スキャン

トル