

プロジェクタを用いた情報投影による 印刷文書へのインタラクティブ性の付加

鈴木 若菜[†] 竹田一貴^{††} 外山 託海^{†††} 黄瀬 浩一^{††}

[†] 大阪府立大学工学部

^{††} 大阪府立大学大学院工学研究科

〒 599-8531 大阪府堺市中区学園町 1-1

^{†††} ドイツ人工知能研究センター (DFKI)

Trippstadter Strasse 122 D-67663 Kaiserslautern, Germany

E-mail: [†]{wakana,takeda}@m.cs.osakafu-u.ac.jp, ^{††}Takumi.Toyama@dfki.uni-kl.de,

^{†††}kise@cs.osakafu-u.ac.jp

あらまし 本稿では、印刷文書に対してプロジェクタで情報を投影し、付加するシステムを提案する。本システムでは、まず、Web カメラで印刷文書を撮影し文書画像を検索する。これにより、情報投影の対象となる文書を特定すると共に、カメラ上での文書の位置を特定する。文書位置の特定により実際の印刷文書上の意図したところへ情報を投影することができる。また、画像処理によってユーザの操作を認識し、それに応じた更なる情報をプロジェクタにより提示する。このシステムを利用することで、電子文書、印刷文書にかかわらず同一のサービスを提供することができるようになる。

キーワード モバイルプロジェクタ、プロジェクタ・カメラ、LLAH

1. はじめに

近年、タブレット端末の高性能化と急速な普及に伴い、電子文書に対する需要が増加している。電子文書には、文書中に情報を埋め込んだり、その関連付けをしたりすることができるという利点がある。例えば、リンクが張られた図や写真を選択すると、リンク先に移動し、元の文書にはない情報を得ることができる。このような電子文書の特徴を生かしたサービスの例として、Layered Reading [1] や Qlippy [2] などが挙げられる。Layered Reading は電子書籍上にレイヤーを乗せ、そのレイヤーに書籍の付加情報を書き込むという形態をとっている。具体的には、本文を読みながら注釈や解説を表示したり、書面のレイヤー上での友人・知人とのコミュニケーションを可能にしたりする。また、Qlippy では、電子書籍のビューワー機能、コメントとハイライト、記事クリップができ、さらには、それらをウェブ上でソーシャルに共有させることができる。

一方で、新聞や書籍といった紙媒体にも依然として需要はある。電子文書に比べて、メモや書き込みがしやすいといったユーザビリティの高さから印刷文書の必要性も滞ることはない。そこで、電子文書と同じサービスが印刷文書に対しても提供することができれば、人と文書とのかかわり方を大きく変えることができると考えられる。ところが、印刷文書そのものに後から情報を加えることは容易ではない。したがって、印刷文書に対して電子文書のようなサービスを実現しようとする、情報を付加することのできるデバイスを用いて、働きかけをするこ



図 1 モバイルプロジェクタ

とが必要不可欠となる。

外部から情報を付加する方法の1つとして、カメラとプロジェクタを利用した情報の投影が考えられる。カメラとプロ

ジェクタを組み合わせた情報投影技術は、これまでにいくつか提案されている [3] [4] [5] [6]。これは、カメラで撮影した物体を認識し、認識結果によって異なる情報をプロジェクタで投影するという手法である。近年、図 1 に示すようにプロジェクタの小型化が著しく、手のひらに収まる程度の小さなサイズのモバイルプロジェクタが次々と市販されている。プロジェクタの小型化及び低価格化により、手軽に持ち運ぶことができ、需要も増えていくと予想される。したがって、プロジェクタを用いたアプリケーション [7] [8] の有用性もますます高まっていくと考えられる。

この技術を利用することで、印刷文書への情報付加が可能となる。そこで、モバイルプロジェクタを用いた印刷文書へのインタラクティブな情報付加を考える。これは、印刷文書に情報を投影するだけでなく、Web カメラで撮影したユーザの行動を認識し、それを投影する情報にフィードバックさせるというものである。この技術が実現できれば、電子書籍のようなサービスを印刷文書にも提供できるようになる。このような形態の情報投影技術に、[9] [10] などがある。しかし、これらの手法では文書の特定に時間がかかり、大規模なデータベースを扱うことができないという問題がある。本研究では、この問題を解決するために、文書の特定に Locally Likely Arrangement Hashing (LLAH) [11] [12] という文書画像検索法を用いる。LLAH は高いスケーラビリティを持ち、リアルタイム検索を実現できるほどの高速処理が可能であるという特徴を持つ。また、カメラ上での文書の位置を特定できるため、情報を投影する位置をリアルタイムで変更可能であるという利点を持つ。

本稿では、モバイルプロジェクタを用いたインタラクティブな情報投影技術を提案する。インタラクティブなシステムを構築するためには、文書の特定を高速に行う必要がある。これには LLAH を利用する。また、意図した位置に像を投影するために、カメラとプロジェクタの位置関係を利用した幾何変換を用いる。これは、カメラ上での文書の位置をプロジェクタ上の位置に変換することを意味する。そして、ユーザがプロジェクタの投影像を操作できるようにするため、カメラでの撮影画像に対して画像処理を施し、画像認識をする。これにより、プロジェクタの投影像にユーザが働きかけることでユーザに応じたさらなる情報の付加をするシステムを構築することができる。基礎的な実験を行ったところ、文書画像検索を開始してからプロジェクタからの投影像を得るまでの処理時間は、220[ms] であった。

2. システム

2.1 システムの概要

提案システムで用いる装置は、図 2 に示すように、モバイルプロジェクタと Web カメラを組み合わせたもの、そして PC で構成される。

本システムの動作の流れを図 3 に示す。あらかじめ測定したプロジェクタとカメラとの位置関係及び、カメラと文書との位置関係の初期値を用いてキャリブレーションをしておく。まず、Web カメラで印刷文書を撮影し、その画像を検索質問として



図 2 装置

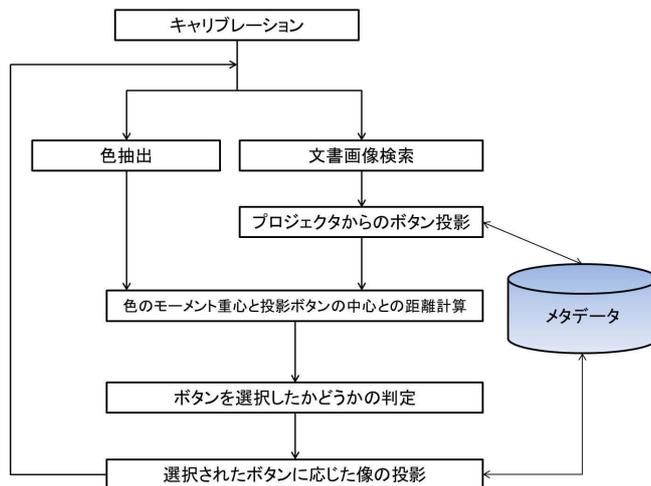


図 3 システムの流れ

文書画像検索を施す。次に、検索結果として得られた文書に対応するメタデータを読み込み、実際の印刷文書上にプロジェクタからの像を投影する。メタデータには、文書名、文書ごとのボタンの投影位置及びボタンが選択された時の各ボタンに対応する投影内容が記述されている。投影された像はボタンの役割を担っており、ユーザが選択できる。ボタンを選択する道具としては、ペンや棒、またはレーザーポインタなどを想定している。ペンや棒の先端、レーザー光は、ある特定の色を持っているとする。撮影画像から色を抽出し、抽出された色のモーメント重心と投影されているボタンの中心との距離計算をすることで、ボタンを選択したかどうかを判定する。その後、選択に応じた像を投影することで印刷文書に対して情報を付加する。これらの処理を繰り返すことでシステムは動作する。それぞれの処理について以下で詳しく説明する。

2.2 キャリブレーション

印刷文書上の意図した位置に像を投影するために、カメラと



図 4 プロジェクタとカメラの距離

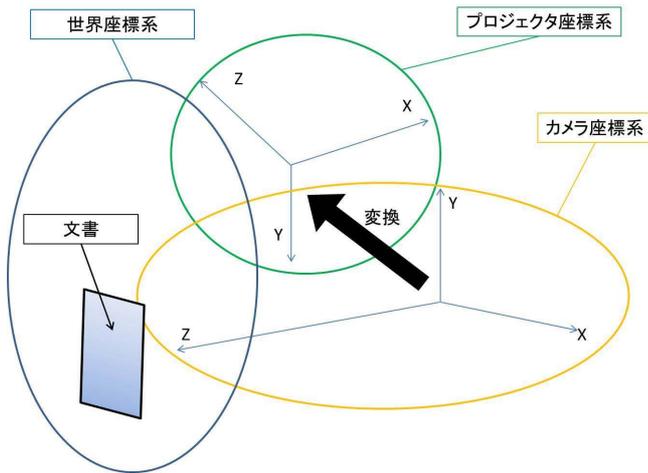


図 5 キャリブレーションにおける座標系の変換

プロジェクタのキャリブレーションが必要となる。このキャリブレーション法について説明する。前提として、プロジェクタとカメラが固定されていて、これらの位置関係が変化しないものとする。まず、図 4 に示すように、カメラのレンズ中心を原点として、カメラとプロジェクタとの X, Y, Z 座標における距離を測定する。測定した X, Y, Z 座標のそれぞれの測定距離および成す角を用いて、並進行列及び回転行列を算出し、変換行列を求める。この変換行列を用いて、図 5 に示すように、カメラ座標をプロジェクタ座標に変換する。これにより、ある文書平面上におけるカメラとプロジェクタの座標の位置を合わせることができ、カメラ座標の任意の位置にプロジェクタの像を投影できるようになる。

2.3 LLAH による文書画像検索

本稿では、LLAH によって撮影している印刷文書を特定する。

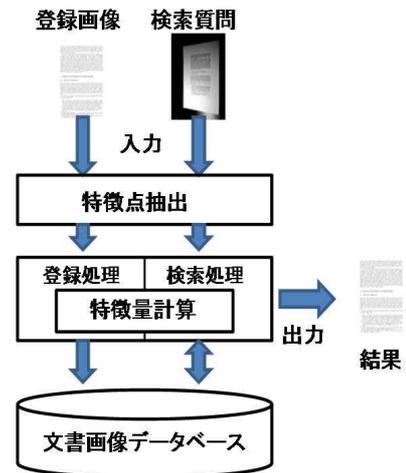


図 6 LLAH の処理の概要

LLAH は、射影歪みや隠れ、紙面の非線形な湾曲という、デジタルカメラで文書を撮影するときに生じる問題に対してロバストであり、リアルタイム検索が可能であるという特徴がある。加えて、さまざまな言語で書かれた文書画像のリアルタイム検索が可能であるという特徴もある。

2.3.1 処理の概要

図 6 に処理の概要を示す。まず、特徴点抽出処理で文書画像は特徴点の集合に変換される。次に、特徴点は登録処理および検索処理に入力される。これらの処理は特徴量計算処理を共有している。登録処理では、各特徴点は独立に、その特徴量に基づいて文書画像データベースに登録される。検索処理では、検索質問の特徴量を用いて文書画像データベースにアクセスし、投票処理で対応する文書画像を決定する。以下では各処理について説明する。

2.3.2 特徴点の抽出処理と特徴量計算

LLAH では特徴点として単語領域の重心を用いる。特徴量は特徴点を表す値であり、LLAH では各特徴点に対する特徴量を求めることで、ロバスト性を実現している。ある特徴点 s の特徴量には、点 s の近傍 4 点 $ABCD$ から、以下の式で求められる値を用いる。

$$\frac{P(A, C, D)}{P(A, B, C)} \quad (1)$$

$P(A, B, C)$ は頂点 ABC からなる三角形の面積である。式 (1) はアフィン不変量であり、局所領域で射影不変量の代わりに使用できる値である。さらに識別性を上げるため、特徴量に近傍点の面積の順位を加える。

2.3.3 登録・検索

LLAH ではハッシュ表を用いて文書画像を登録・検索する。登録処理では、各特徴点はその特徴量を基に、文書 ID、点 ID、特徴量の組をハッシュ表に登録する。検索処理では、検索質問画像から得られた特徴点に対して、特徴量を基にハッシュ表からリストを得る。リストの各項目について特徴量が一致するか調べ、一致していれば登録されている文書 ID を用いて、対応する文書に投票する。各特徴点に対してこれらの処理をし、最終的に得票数が最大となる文書を検索結果とする。文書画像検

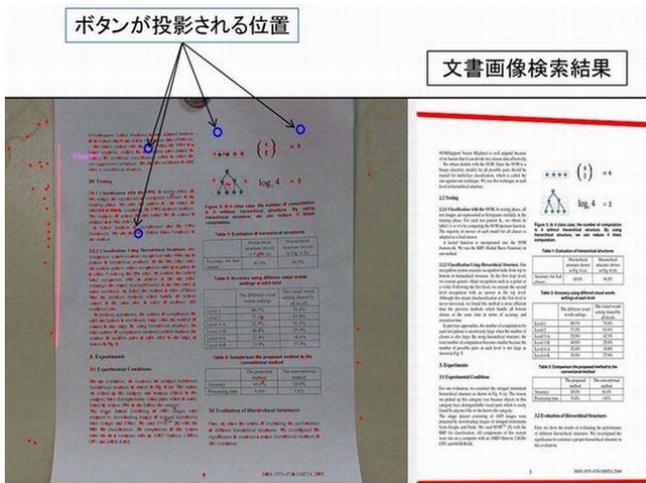


図 7 文書画像検索の様子

索の様子を図 7 に示す。

2.4 カメラ上での文書位置の特定

本システムでは、文書の位置が移動しても、文書上の同一の位置にプロジェクタの像を投影するために、カメラで撮影した画像上のどこに文書が位置しているかを特定する必要がある。ここでは、前提として、文書の位置の移動はキャリブレーション面と同一平面内とする。カメラ上の文書の位置を特定することで、文書の移動に追従してプロジェクタの像を投影することができる。以下でその方法を説明する。LLAH では、検索質問画像とそれに対応する登録画像の特徴点の対応関係を求めることができる。この対応関係を用いて、登録画像から検索質問画像への射影変換パラメータを推定することができる。このパラメータを用いて登録画像の 4 頂点を検索質問画像へ変換することで、カメラ上での文書位置を特定することができる。この位置座標をキャリブレーションにより求まる変換行列を用いて、プロジェクタ座標に変換することで、常に文書上の指定した位置にプロジェクタの像が投影される。

2.5 カメラと文書との距離の算出

プロジェクタとその像の投影面との距離が変化すると、それに伴って像の投影範囲が変化する。投影範囲が変化すると、それに伴い投影像の位置も変化してしまう。そこで、プロジェクタとその投影面である印刷文書との距離が変化しても常に文書上の同一の位置にプロジェクタの像を投影するために、LLAH の射影変換行列を利用して、文書とカメラの距離を算出する。

2.6 選択したボタンを特定するための色抽出

本システムでは、文書上に投影されたボタンをユーザがペンで指すと、その選択した情報を得ることができる。ユーザが文書上のどのボタンを指したかを特定するために、ある特定の色の色抽出を利用する。本システムではボタンを指す道具に先端が緑色のものを用いる。以下に緑色の色抽出の処理内容を述べる。まず、カメラからの入力画像の RGB 系色空間を R, G, B に分解する。分解された画像のうちペン先の色の要素、つまり分解された画像のうちの G の要素が 100 以上で、R と B の要素より 1.5 倍以上あるピクセルを抽出する。緑色の色抽出の様

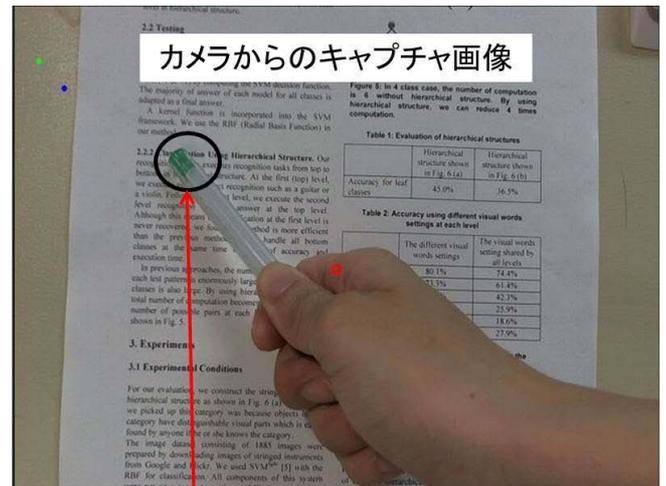


図 8 色抽出

子を図 8 に示す。

2.7 ボタン選択の判定法

ユーザがプロジェクタで投影されたボタンをペンで選択したかどうかの判定方法について説明する。判定には 2 つの条件を用いる。第 1 は、ペン先とボタンとの距離による条件である。色抽出によって抽出したペン先の領域のモーメント重心を計算する。モーメント重心と投影されているボタンすべての中心との距離をそれぞれ計算し、距離が最小となるボタンをユーザが選択したボタンの候補とする。そして、モーメント重心と候補となったボタンの中心との距離が閾値以下であればそのボタンを指したと判定する。しかしこの条件だけでは、ボタンの上を通過するだけという場合でもボタンを選択したと判定されてしまう。このような誤判定を防ぐため、第 2 の条件として、時間の条件を用いる。最少距離で閾値以下であるボタンが一定フレーム以上同一であるという条件を満たしたときのみそのボタンを指したと判定することとする。ユーザの選択候補と判定した場合、プロジェクタの像により候補ボタンを枠で囲むというフィードバック機能を備えている。ペンのタッチの様子を図 9 に示す。ここでは、LLAH の検索結果により得られた文書名と選択候補ボタンのフィードバック機能により候補ボタンが枠で囲まれた像が文書上に投影されている。また、選択された情報が投影される様子を図 10 に示す。これは、ボタンを選択したこ

とで、文書中のキーワードの説明を投影している様子である。

2.8 アプリケーション

本研究では、デバイスに依存しないアプリケーションを実現することが最大の目的である。本稿では、その第一段階として、印刷文書を対象としている。文書であれば、印刷文書をわざわざPCなどに取り込み使用する手間がなく、印刷文書そのものがPCのディスプレイやスマートフォンなどのデバイスの機能を担うことができる。

LLAHによる文書画像検索を用いることで多数の言語の多数の文書が瞬時に検索可能となり、プロジェクタとカメラを駆使することでその文書に埋め込まれている情報を必要な時に必要な分だけ容易に取り出すことができる。具体的には、例えば、英語で書かれた文書については、文書中のキーワードの説明が用意されていれば、そのキーワードを選択すると説明を得ることができる。また、わからない単語が生じればその訳語を提示したり、参照や付録などの記述があれば、その参照先の情報を提示したりすることも可能となる。

3. 実験

本稿で提案したシステムでは、ユーザビリティの観点から、ユーザの起こすアクションをリアルタイムで認識する必要がある。また、文書の特定の位置にプロジェクタの像を投影することも必要である。このためには、カメラ上からプロジェクタへ正確に座標変換されなければならない。そこで、実験により、本システムの処理時間とカメラからプロジェクタへの座標変換による投影位置の精度について調べた。

実験で用いたデータベースの登録ページ数は、10,000である。データベースから10枚の文書画像を抽出し、印刷して検索質問とした。Webカメラの解像度は1280×960である。なお、実験に用いた計算機は、CPU 2.8GHz、メモリ 8GBのものである。

3.1 処理時間

本システムの色抽出及びLLAHにかかる処理時間を測定した。実験の際にはWebカメラを印刷文書から距離20cmの場所に固定して撮影した。ここで、処理時間とは、Webカメラのキャプチャ画像の処理時間をベースとして、その処理のループの中で、色抽出や文書画像検索、プロジェクタの像の投影などの処理が終了しているかを調べてそれを計測している。LLAHにおける平均処理時間は、色抽出が62ms、LLAHが220msであった。これは、ユーザが反応を待てる時間であると考えられる。したがって、本システムはリアルタイム性を保持していると考えられる。

3.2 カメラからプロジェクタへの変換精度

次に、距離変化による安定性を確認するため、カメラ上での投影位置とそれを変換した実際の投影像の位置との誤差を調べた。実験では文書とカメラの距離を図11のように、200mm、250mm、300mm離し、それぞれの距離において上下左右に50cmずつずらした時の誤差を測定した。カメラと文書の距離が300mmで中央のときの様子を図12に、得られた結果を表1に示す。場所によって精度にばらつきが生じていることがわか

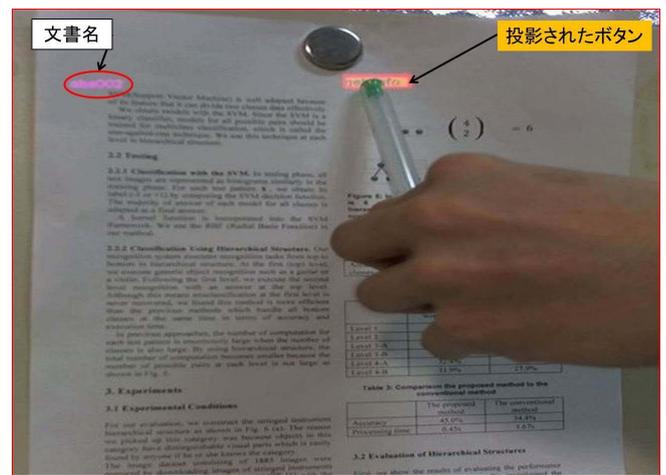


図9 ボタンタッチの様子

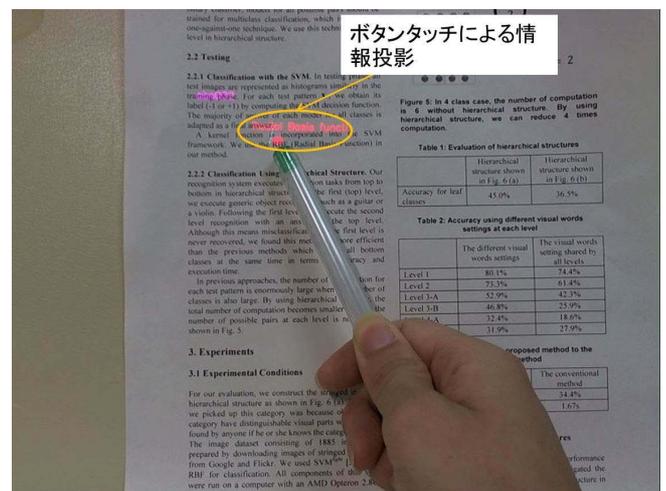


図10 ボタンタッチによる投影像

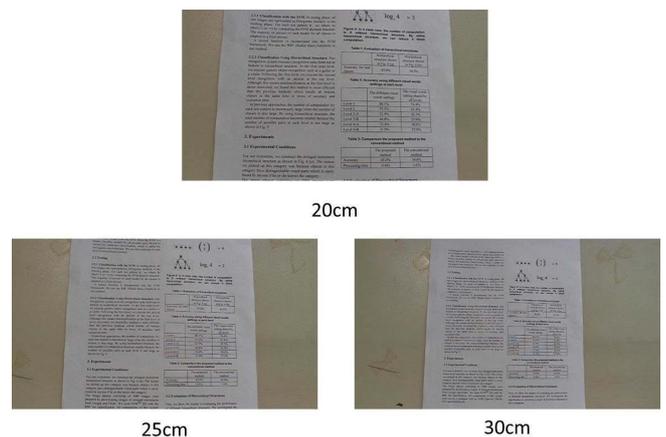


図11 距離ごとの文書画像

る。図13に示すように格子状の像を投影すると、端に行くにつれ直線が歪んでいることがわかる。このことから、プロジェクタの像自体に歪みがあり、これが原因の一つと考えられる。しかし、誤差は全て10mm以内に収まっており、一定の精度で変換できていると考えられる。

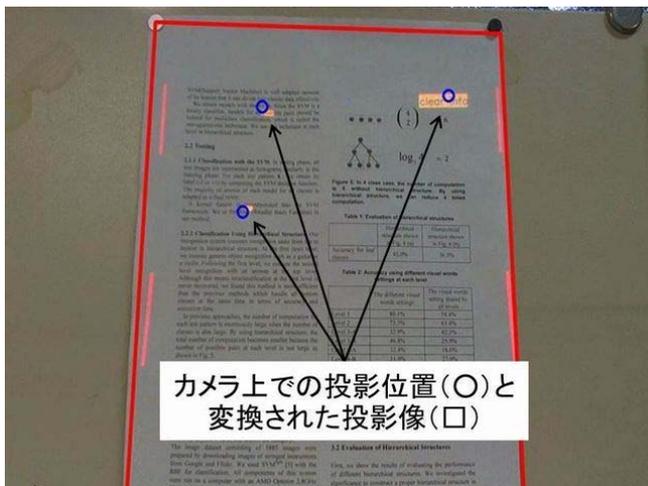


図 12 投影位置と投影像

3.3 色抽出の検討

色抽出に関して、環境光の影響により濃い緑色が抽出されなかったり、また、逆に緑色ではないのに誤って抽出されたりしたものが観察された。このことにより、選択していないボタンが反応してしまい誤作動を引き起こしてしまっていた。したがって、今後の課題として、ボタン選択の手段や認識方法の改善が挙げられる。

4. ま と め

本研究では、印刷文書に対して電子文書のようなサービスを実現するためのプロジェクタとカメラを用いたシステムを提案した。ペンをシステムの操作の手段とし、撮影画像に対して色抽出処理をすることで、プロジェクタからの投影像のタッチ判定をした。そして、印刷文書に対してユーザの選択に応じた情報を付加することができた。したがって、あらゆる文書に対して固有の情報を用意することが可能である。また、LLAH による文書画像検索からカメラ上での文書位置が特定できることから、文書を固定せず動かしても文書の特定の場所にプロジェクタの像を投影することができた。

本稿ではテキストが主体の文書を対象としていたが、このような文書だけでなく絵本や図鑑などの紙媒体を対象とすることも必要である。これにより、幅広い層のユーザが利用できるようになる。また、プロジェクタだけでなく、アイトラッカーなどのデバイスとも組み合わせることでさらなるサービスの幅が広がると考えられる。本稿で用いた装置は、机の上などにおいて使用する固定型の装置である。今後、カメラおよびプロジェクタのさらなる小型化が実現されると予想されるため、ウェアラブルな装置およびそれに対応したシステムを構築することも課題の一つである。

謝辞 本研究の一部は、JST CREST および日本学術振興会科学研究費補助金基盤研究 (B) (22300062)、挑戦的萌芽研究 (21650026) の補助による。

文 献

- [1] <http://84dialog.blogspot.com/2010/03/layered-reading.html>
 [2] <http://jp.techcrunch.com/archives/jp-20100514-qlippy-suggest->

表 1 カメラからプロジェクタへの変換誤差

文書とカメラの距離 [mm]	200	250	300
中央の時 [mm]	5.1	6.8	3.8
上に 5cm の時 [mm]	4.8	7.5	3.7
下に 5cm の時 [mm]	5.7	7.5	4.5
右に 5cm の時 [mm]	1.7	4.0	5.7
左に 5cm の時 [mm]	5.3	3.3	4.8

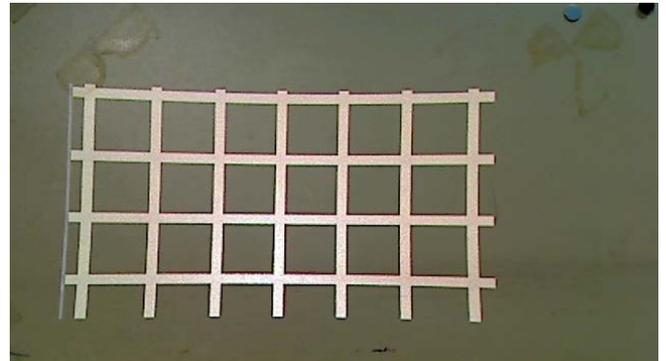


図 13 プロジェクタの投影像

- you-a-new-reading-experience-by-using-a-social-power/
 [3] H.S.T. Grossman, G. Fitzmaurice, F. Guimbretiere, A. Khan, R. Attar, and G. Kurtenbach, "Penlight: Combining a mobile projector and a digital pen," CHI 2009 New Tabletop Input and Output Methods, pp.143-152, apr 2009.
 [4] S. Audet, M. Okutomi, and M. Tanaka, "Direct image alignment of projector-camera systems with planar surfaces," Proceedings of The 23rd IEEE Conference on Computer Vision and Pattern Recognition, pp.303-310, jun 2010.
 [5] 天野敏之, 加藤博一, "プロジェクタカメラフィードバック系によるアビランス強調," 画像の認識・理解シンポジウム (MIRU2008), pp.79-86, jul 2008.
 [6] 天野敏之, 加藤博一, "モデル予測制御を用いたプロジェクタカメラ系によるアビランス制御," 画像の認識・理解シンポジウム (MIRU2010), pp.746-753, jul 2010.
 [7] <http://research.microsoft.com/en-us/news/features/touch-101711.aspx>
 [8] <http://japan.internet.com/webtech/20111019/3.html>
 [9] Q. Liu and C. Liao, "PaperUI," CBDAR2011, sep 2011.
 [10] C. Liao, H. Tang, Q. Liu, P. Chiu, and F. Chen, "Fact: Fine-grained cross-media interaction with documents via a portable hybrid paper-laptop interface," pp.361-370, 2011.
 [11] 中居友弘, 黄瀬浩一, 岩村雅一, "特徴点の局所的配置に基づくリアルタイム文書画像検索とその拡張現実への応用," 電子情報通信学会技術研究報告, vol.106, no.PRMU-229, pp.41-48, sep 2006.
 [12] 竹田一貴, 黄瀬浩一, 岩村雅一, "1,000万ページのデータベースを対象とした実時間文書画像検索のためのメモリ削減と安定性向上," 電子情報通信学会技術研究報告, vol.110, no.467, mar 2011.