

検索対象としてのデスクトップイメージ画像

Desktop Images as Target of Information Retrieval

梅村恭司¹ 武並佳則² 折原幸治² 熊谷摩美子¹ 杉浦遼一¹ 若松翔¹ 月田晴貴¹
Kyoji Umemura¹, Yoshinori Takenami², Orihara Koji², Mamiko Kumagai¹, Roichi Sugiura¹,
Sho Wkamatsu¹ and Tukita Haruki¹

¹豊橋技術科学大学 情報・知能工学系

¹Toyohashi University of Technology, Dept. Computer Science

²住友電気情報システム ビジネスソリューション事業本部

²Sumitomo Electric Information Systems Co., Ltd.

Abstract: Recently, major information for human comes and go through computer, and recording the activity on computer can be regarded as a kind of life log of human. Unlike life log in video form, the life log of computer records are rich in text information. This make it possible to handle the record more effectively than video. This report presents a prototype system for retrieval the captured image of desk top of computer. This report also discusses various characteristic of captured image collection compared with text collection. Finally, this report presents some situation where the proposed system will be useful.

概要

人間の活動に関わる検索を考えたときに、コンピュータ画面を通じて情報を取り入れる比率は大きくなっており、人間の活動記録の検索の一つの例としてデスクトップイメージの検索を考えることができる。一方で、一般的なビデオ画像に比べて画像のなかに文字が多く含まれ、それによって検索できる利点がある。われわれは、これを具体的に示すために、検索のプロトタイプを作成している。本発表では、コンピュータの操作画面画像（デスクトップイメージ画像）の検索のプロトタイプシステムを通じて、検索対象としての特色と検索の意義を具体的に報告する。

1. はじめに

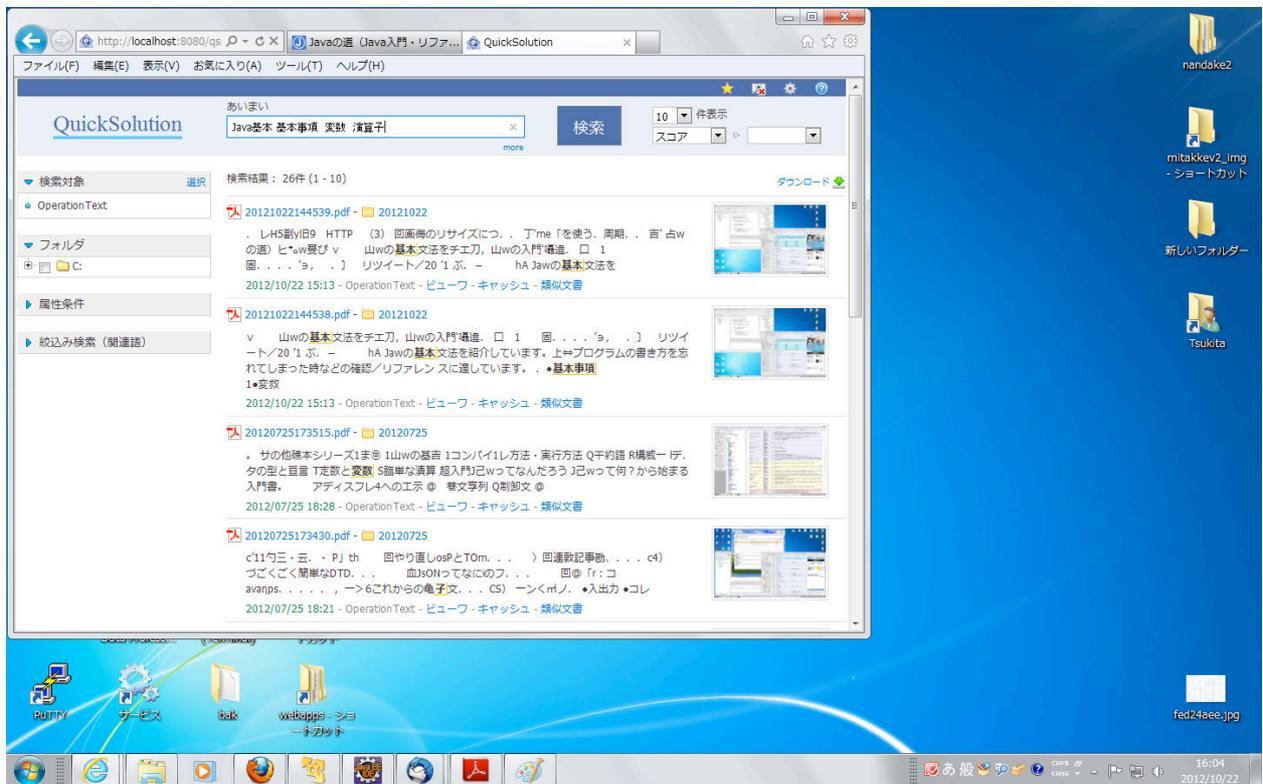
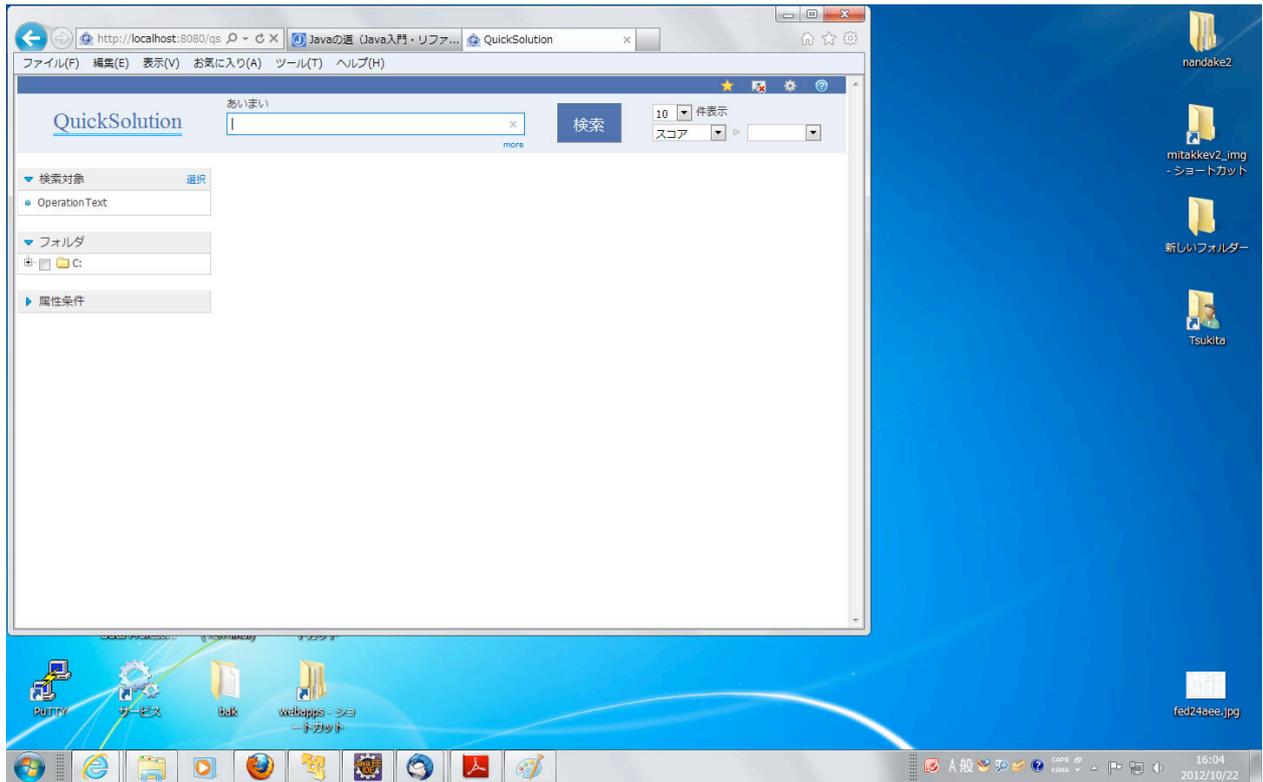
ポータブルデバイスとネットワーク発展により、人間が取得する全ての情報を記録して、活用することに注目が集まっている[1]。ここにおいて取得されるものは、人間の目がとらえる映像が主たる要素を占める。この情報は先頭から順番に調べていくのには大き過ぎるものであるので、検索のシステムが必要であるが、映像の検索には課題が多い。

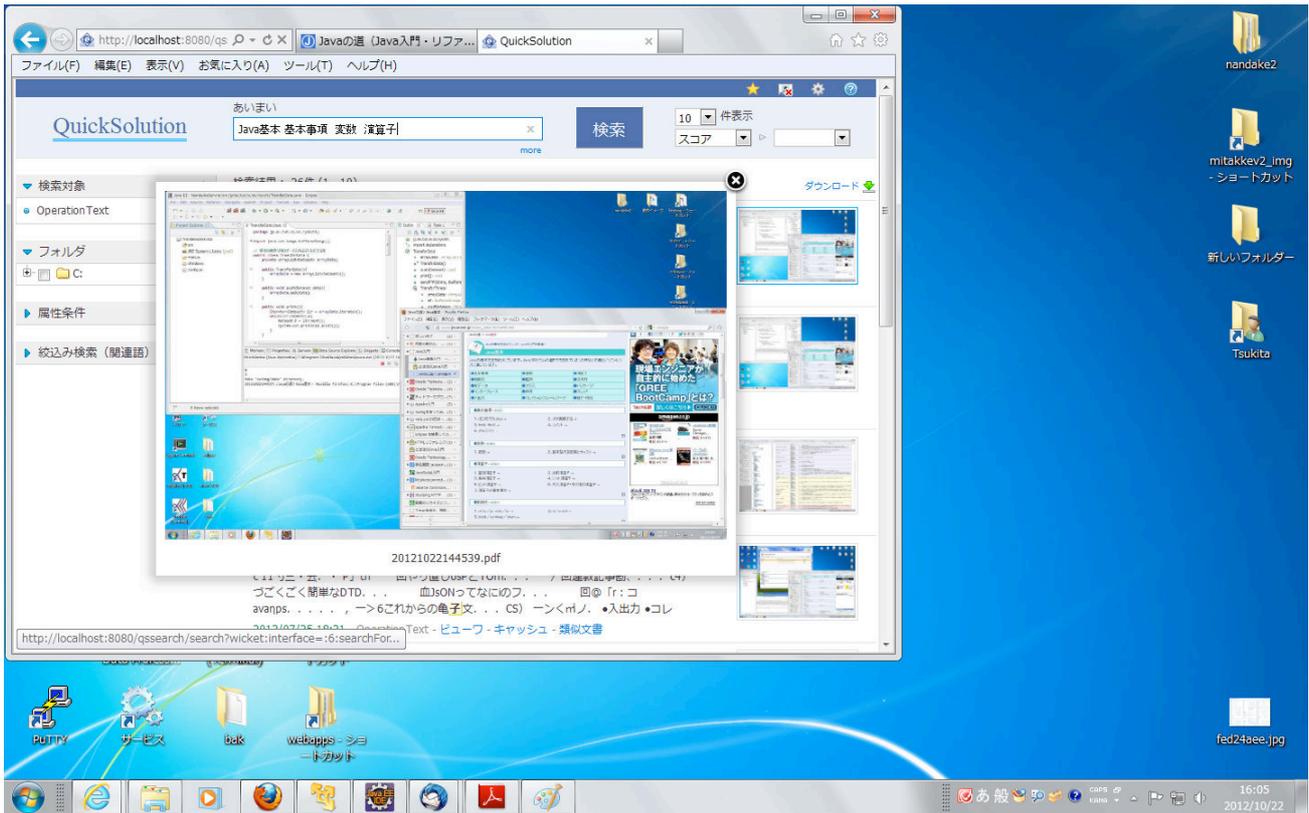
近年、書籍、テレビ、新聞、会議などの多くの情報源がコンピュータを経由して利用できるようになり、その比率は大きくなる傾向がある。ここで、装

着型の記録デバイスの代わりに、コンピュータの画面を記録することを考えた。コンピュータ経由以外の人間の情報活動は対象にならないかわりに、携帯デバイスを装着、管理する人への負担がなくなると同時に、画像としては鮮明な画像が得られ、また、画像のなかにも索引の対象となりうる文字が多く含まれる特徴がある。

コンピュータの活動を記録し検索する提案はなされている[2]が、ここでの対象は利用したメールやドキュメントそのものであり、それが画面に現れている画像とは異なっている。デスクトップ画像であれば、ある文章を入力していたとき、その編集のイメージだけでなく、参考としてひらいていた Web ブラウザのイメージが同時に検索できる。このような「同時にひらいている」という情報は、ユーザの経験を構成する重要な要素であるが、メールやドキュメントだけを対象とすると、その情報が失われてしまう。

このような背景から、デスクトップイメージの検索を考えることができる。一方で、一般的なビデオ画像に比べて画像のなかに文字が多く含まれ、それによって検索できる利点をもちつつ、画像という人間が見た状況を検索対象としているという特徴をもつ。本稿では、これを具体的に示すために、検索のプロトタイプを作成し判明した知見を報告する。





2. プロトタイプシステムの操作手順

まず、デスクトップ画像はタイマーによって定期的に（プロトタイプシステムでは 30 秒に 1 回、設定で変更可能）検索サーバに転送しておく。サーバは、それを定期的（プロトタイプシステムでは、1 日に 1 回、設定で変更可能）に索引付けをする。

その準備が済んでいる状態でのプロトタイプシステムの操作状態を図 1 から図 3 に示す。図 1 においては、検索のために検索サーバの検索要求用の url を Web ブラウザでひらいたところである。これは市販の検索システム[5]の検索画面、そのままである。

ここから、検索語として、「Java 基本 基本事項 変数 演算子」を入力し、検索ボタンをマウスでクリックすると図 2 の画面になる。検索の意図は、以前に Java の言語機能をまとめてある Web ページをみたことがあるので、そのときのデスクトップイメージを想定している。図 2 においては、検索画面に含まれている文字列をスニペットとし、検索画面全体を小さくしたサムネイルがリストで表示される。画面は、1 日毎に日付に相当するフォルダに分かれているので、そのフォルダの情報と、時間情報に従ったファイル名の検索対象が表示される。検索のオ

プションで、時間指定もできるほか、ファイル名に日付と時間があることを利用して、検索入力にその時間入れることもできる。

サムネイルで、目的のものと思われる画面があった場合には、その画面のさらに拡大したイメージを見る事ができる。その操作を行っているのが図 3 である。これが正しいということにあれば、Acrobat Reader などの PDF 文書と関連つけられたアプリケーションで PDF に変換されたデスクトップ画像をひらくことができ、印刷やクリップボードでの利用ができる。

3. プロトタイプシステムの構造

プロトタイプシステムの構造は、すでに既に発表したシステム[3]と同一であるが、検索エンジン[5]の機能強化のために、既に発表のシステムにくらべ実用的なものになっている。

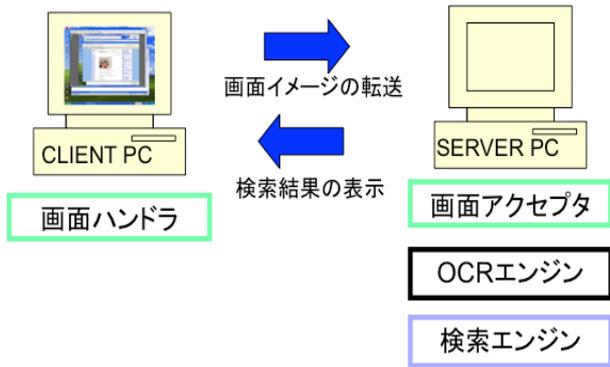


図4 プロトタイプシステムの構造

プロトタイプシステムは検索サーバとクライアントからなるサーバクライアントモデルで作成した。一つの検索サーバで複数のクライアントのデスクトップの検索が行える。図4は、[3]と同一のものであるが、サーバとクライアントに配置されているコンポーネントの名前を示す。以下に、それぞれのコンポーネントの機能を説明する。

画面ハンドラは、設定された一定間隔ごとに画面をバックグラウンドでキャプチャし、それをサーバに転送する。クライアントに追加するコンポーネントであるので、処理は最小限にした。画面をキャプチャして、サーバに転送する程度であれば、マルチコアが一般的となった近年のコンピュータでは、特に存在を感じさせない程度の負荷のコンポーネントである。

画面アクセプタは、ネットワークから画像イメージを取得したのち、相手のIPアドレスと日付から適切な格納ディレクトリを決め、それを画像jpeg形式で格納する。

OCRエンジンは、設定されたディレクトリ/フォルダに格納された画像ファイル进行处理し、同じ名前のPDFファイルに変換する。そのときに、OCR処理した文字列は透明のテキストとして、結果のPDFファイルに埋め込む動作をする。ここは、市販のOCRソフト[4]を使っている。

OCRで処理されたデスクトップイメージは、コンピュータのプログラムからするとテキストを含むPDFファイルの形式であり、処理をするときには通常のテキストのPDFファイルと同様に操作できる。それゆえ、PDFファイルを生成するフレームワークを作成したことで、市販の検索システムに手を加えることなく、プロトタイプシステムが動作した。

検索エンジンは、設定されたディレクトリ/フォルダにあるPDFを検索対象にし、一定時間ごとに索引を更新する。そして、Webブラウザを経由して、検索質問を受付、Webブラウザで結果を返す。この

検索システムは製品[5]をそのまま使用している。この応用のためにカスタマイズは行っていない。

4. OCR アプローチ

デスクトップイメージ画像には文字情報が含まれることが多いということが本システムの着眼点であるが、この文字情報を画像から取り出す操作が必要である。

もともと、デスクトップに文字を書くときには、もとの文字列をコンピュータが表示しているのも、そのレベルでのオペレーティングシステムから情報を取り出すというアプローチも考えられるが、われわれは、ここにOCRを使うことにした。

OCRはコンピュータに負荷のかかる操作であり、また、読み取った結果に誤りが含まれる率も高くなる。正確な検索のためには、OCRを利用することは問題が生じるかもしれないのは事実である。

しかしながら、Webブラウザで表示されている情報、たとえば、スクリーンショットの一部を利用した操作説明などでは、人間に読める文字であってもコンピュータは文字出力でないこともある。また、OSの機能を使おうとするために、OSのバージョンの違いに影響を受けやすかったり、複数のOSで画面ハンドラを作成するのが難しくなったりする。

以上、まとめると、OCRを利用することで、処理装置への負荷や検索精度の低下が心配されるが、一方で、文字が取り出せる範囲が広がることと、プロトタイプシステム全体の機種依存性が下がることとなる。

この応用では、OCRで読まれた文字列について検索をかけるので、表記の誤りに強い検索エンジンが必要である。使用した検索エンジンは、文字バイグラムをベースに検索する機能があり、キーワードの一部がOCRの読み取りの誤りで失われても、検索はできるという性質があり、OCRの誤りで検索精度が低下する可能性はあるが、まったく検索できないということはない。

5. サムネイル機能

初期のプロトタイプシステム[3]で使用したシステムと現在のシステムとを比較すると、検索を実行した直後の図2で現れるウィンドウのサムネイル機能の有無が大きな違いとなる。

この機能は、検索エンジンの機能のバージョンアップによるものである。画面ハンドラや画面アクセプタについて本質的に[3]で作成したものと同一であり、OCRもバージョンアップにより読み取り精度

が向上したものの、ユーザのインタフェースという観点からは OCR による変更はない。

サムネイルは、検索対象を選ぶときに、スニペットのテキストを補佐し、全体的なレイアウトから対象を選べる機能として一般的なものであるが、自分が過去に見たデスクトップ画面を検索するときには、ウィンドウの色やレイアウトがシーンを選ぶときのとりかかりとして、特に有用な機能となる。

6. ビューワ機能

初期のプロトタイプシステムと比較して、もう一つ重要な検索エンジンの機能拡張はビューワ機能である。ビューワ機能は、サムネイルで見込みがありそうなデスクトップ画像について、検索ワードをハイライトした状態で、内容を確認できる機能である。

どのような画面化を確認することは、デスクトップ画像を検索するときの主要な目的と考えられ、ビューワで検索ワードがどこに表示されるかがわかることによって操作性の改善がなされた。

7. システムの負荷

現状のシステムはコンセプトの実現性の検討が目的であり、プロトタイプであって、製品ではないのでシステムの負荷は第一義ときには考慮せずに設計したが、大規模な実験にも耐えるシステムになっている。

まず、クライアントシステムについては、操作画面をネットワークに送る操作であり、その間隔として、画面がかなり変化するであろうという時間間隔、具体的には数秒から数十秒の間を想定しており、操作の妨げになるような負荷ではないし、実際に画面ハンドラが動作していることをコンピュータ操作で意識することはない。

処理速度のネックはコンピュータ画像の OCR 処理である。この処理は、ディレクトリの監視などのタイミングもあるが、1画面に十数秒かかる。しかしながら、OCR 処理は画像取得に間に合うようにおこなう必要はなく、操作がなにもないときに、一度に行うことができる。また、将来、OCR 処理がネックになったとしても画像毎に独立の処理であるので、必要に応じて、複数のコンピュータで分担する方法も容易に実装できると考えられる。

プロトタイプシステムでは固定の時間間隔でデスクトップ画像を転送するため、ハードディスク領域を圧迫することが心配されるが、消費するディスク容量を計算すると、実験のためのプロトタイプシ

テムとしては、大規模な実験にも耐えられるものであることがわかる。具体的に、1日8時間、1000日の記録をとることを考える。10秒に1枚、デスクトップ画像を取得するとすると、1分に6枚、1時間に360枚、1日に8,640枚、これを約10,000枚と考えると、100日で1,000,000枚となる。1920×1080ドットの解像度の画面から生成した一枚のPDF画像は、200キロバイト程度なので、これは、200ギガバイトのファイル容量で保持できる量となる。

将来的には、デスクトップ画像であることを想定した簡略な OCR 処理で OCR の負荷を下げることができそうであるし、ファイル容量についても、ほぼ同じ画像は処理しないなどの工夫ができると考えられる。

8. 多くの類似文書が存在する問題

実際に検索を行ってみると、デスクトップイメージの性質に由来する問題がいくつか生じたので、それについて報告する。まず、この検索対象には類似の文書が数多く存在することがある。

図2において、複数の画面で、ほぼ同一のサムネイルとスニペットが表示されていることが分かる。これは、30秒おきに画面を取得しているためであり、画面上は、まったく変化がなくても新しい画像が新しいドキュメントとして生成される。これは、文書ファイルにおいて、自動セーブ機能により、多くのバージョンが格納されている状況での文書検索に相当する問題といえる。

この性質により、2つの課題が生じる。一つ目は、検索結果を提示するときに、類似している文書(画像)から代表の文書(画像)を選び出し、それを提示して、多様な候補を検索結果候補として選ぶ課題である。通常の見出しでも、同様な課題があり、クラスタリングなどの手法が応用されるが、この場合には、取得時間あるいはバージョンが利用できるため、より精度よく代表の文書(画像)を選ぶ事ができる可能性があるが、それは面白い課題である。

もう一つの課題は、検索における「検索語」に対する重みの問題である。ここで「検索語」と括弧をつけて記述したのは、使用したシステムでは文字バイグラムも「検索語」として扱っており、いわゆる単語でないものも含まれるためである。「検索語」の重みの決定には、文書頻度をもとに多くの式が考案されているが、文書の生成モデルとしては、それぞれの文書が独立の確率モデルから生成されると仮定されることが多く、使用した検索システムも、その仮定にしたがって語の重みを計算して、検索結果の候補の順位を決定している。今回の状況では、前提

となる文書生成の仮定が崩れており、そのもとでは適切な「検索語」の重みをあたりに検討する必要がある、これも重要な課題である。

9. 時間軸の扱いの問題

検索の対象となる PDF ファイルは、定まった時間で取得された画面イメージであるので、全体として操作画面を微速度撮影した映像とみなすことができる。これは、通常の実験とは異なる性質である。通常の実験では、個々は独立のファイルの集合と考えるので、検索結果について、前後を見る機能がない。

このような機能はビデオのフレームを対象とする検索のように、検索結果から前後を表示したり、ビデオのキーフレームに対応する画面を特定したりすることが必要となる。

ここでも、自動保存があるようなアプリケーションにおいて、すべての中間状態を含めて文書検索をする状況とデスクトップイメージの検索は類似性がある。この問題において、デスクトップイメージの検索は1ストリームのデータであるのに対し、文書情報は、ファイル上は1ストリームであっても、カットペーストにより枝分かれやマージが生じうるので、より複雑な検索対象といえる。

以上のことより、バージョンや名前の変更などを含めた文書の時間軸変化を考慮した検索研究を行うための準備として、デスクトップ画像を検索対象にして検索機能の検討をおこなうことは価値があるように思える。

10. 画像検索の実使用のシナリオ

検索における新技術の開発のためだけでなく、実際の状況でも画面検索が有効と考えられるケースがある。

最初のシナリオは、コンピュータを用いた演習を行っているケースであり、教師が生徒ごとの進捗状況をまとめたいというシナリオである。課題の区切りごとに、操作画面に特徴のある文字列、たとえば、課題を遂行するプログラムに含まれる文字列などを手掛かりに、検索システムを利用して進捗を一覧にして、課題の難度を確認することや、進捗に問題のある生徒を発見するというシナリオは存在する。

次のシナリオは、一連の操作をしているときに、エラーがでた状況を検索し、そこから画面を巻き戻して見て、エラーの原因をさぐるようなシナリオである。

どちらの場合も、操作画面をビデオで取得して、早送り、巻き戻しなどで場所を特定するより、キー

ワードを指定して検索するほうが効果があることが期待できる。現状のプロトタイプでもこのようなシナリオのために利用できるが、より操作性を高めるように検討すること、検索システムのユーザインタフェースを改良していくこともできると考えられる。

まとめ

本稿では、デスクトップ画像を検索するという問題に対し、既存の OCR システムと検索システムを利用してプロトタイプの作成をし、そのプロトタイプの操作を報告した。そして、プロトタイプシステムがコンピュータにかける負荷の大きさを示し、現在のコンピュータでは、十分に実験できる対象であることを述べた。

そして、この課題は、自動保存のある文書と類似性があることを述べ、その準備として問題が単純化されており、一方で、ビデオ検索との接点もあることを述べた。

最後に、デスクトップ画像の検索が、それだけで応用の価値があると考えられるシナリオを示し、プロトタイプから次のステップに行くときの方向性を示唆した。

参考文献：

- [1] 堀 鉄郎, 相澤 清晴: “ライフログビデオのためのコンテンツ推定”, 映像情報メディア学会技術報告 Vol. 27, No. 72 pp. 67-72, (2003)
- [2] Susan Dumais: Stuff I've Seen: a system for personal information retrieval and reuse, ACM SIGIR 2003, pp. 72-79, (2003)
- [3] 熊谷摩美子, 梅村恭司, 岡部正幸, 阿部洋丈: 操作画面を対象とする検索システムの構築, 情報処理学会 72 回全国大会, 4R-2, pp767-768, (2010)
- [4] 株式会社ハイパーギア: HG/PscanServPlus 製品ページ, <http://www.hypergear.com/index.html>, (2012)
- [5] 住友電工情報システム株式会社: Quick Solution, <http://www.sei-info.co.jp/quicksolution/index.html>, (2012)