

情報獲得機能を持つウェアラブル・オーサリングシステム

大内 一成 大盛 善啓 松下 宗一郎 土井美和子

(株)東芝 研究開発センター ヒューマンインタフェースラボラトリー

E-mail: {kazushige.ouchi, yoshihiro.ohmori, genso.matsubara, miwako.doi}@toshiba.co.jp

1. はじめに

日常生活において、過去に収集したマルチメディアデータの編集、レポート作成などの作業は非常に手間がかかるのが現状である。これは、撮影した画像データとそれに対するメモなどが、それぞれ関連付けられずに蓄積されていることに起因する。

一方、ウェアラブルコンピュータにおいてストレスのない入力を実現するためには Context-awareness が利用される[1]。Context-awareness を利用し情報提示を行う試みや[2]、データ入力の際に Context-awareness を利用する試み[3]、などが報告されているが、データ入力、情報提示の両方に利用されるべきであると考えられる。

そこで我々は、ユーザの動作を認識する状況認識センサを手首に装着し、ハンズフリーでデータ入力を行い、即座にそれらのマルチメディアデータを構造化して HTML 形式で保存するウェアラブル・オーサリングシステムを提案する。まず、従来のデータ収集システムにおける問題点について触れ、次に我々が開発したウェアラブル・オーサリングシステムについて説明し、展示会で実際に本システムを使用した際の結果を紹介する。

2. 従来のシステム

例えば展示会を見学してレポートを作成する際、見学者はデジタルカメラでたくさんの画像を撮影し、それとは別にメモを書く。そして、オフィスや自宅に戻ってからそれらのデータを整理してレポートを作成する。その際、撮影時の鮮明な印象が失われている場合が多く、また構造化されていないマルチメディアデータの整理には非常に手間がかかるという問題があった。

また、デジタルカメラの中には、撮影した画像に対し、一定時間の音声メモを録音することができるものがあるが、後から音声メモをデータの検索等に利用できるといったものはない。

このように、従来のデータ収集システムは、あくまで素材としてのマルチメディアデータを収集するためだけに作られており、収集されたデータ間の高度な関連付けや、自動編集などには適しておらず、

A Wearable Authoring System Using Organized Multimedia Data
Kazushige Ouchi, Yoshihiro Ohmori, Soichiro Matsushita
and Miwako Doi
Human Interface Laboratory, TOSHIBA Research and Development Center

それらの作業は後から別の編集用システムを利用して行わざるをえなかった。

3. システム概要

この問題点を解決するために、音声認識により認識されたキーワードを検索タグとして利用できるようにマルチメディアデータを構造化するウェアラブル・オーサリングシステムを提案する。図1に、本システムの概観と、構造化されたマルチメディアデータの例を示す。本システムは、状況認識センサ、帽子に取り付けられたカメラ、マイク、ミニノートPC(Libretto ff 1100)で構成される。ユーザの手首に取り付けられた状況認識センサは、加速度センサと MPU で構成される。手首の動作の認識はすべてこの MPU で処理され、PC には認識結果のみが送られる。従って、PC の処理に負荷をかけることなく、動作の認識を行うことが可能である。この状況認識センサにより認識された動作により、画像の撮影、音声メモの録音、音声キーワードの認識、の3つの入力モードをハンズフリーで切り替えることができる。

PC では、マルチメディアデータ構造化と音声認識の処理を行う。音声認識エンジンには弊社で開発した音声認識モジュールを用いてキーワード認識をする。状況認識センサにより音声メモ入力と音声キーワード入力を切り替える。これらの音声データとタイムスタンプは自動的に画像データと関連付けられ[4]、検索タグとして利用することができる。このようにして構造化されたマルチメディアデータは互いにリンクされ、編集しなくても実用に耐えるデータとして HTML 形式で保存される。

このようにして、本システムでは、収集したマルチメディアデータがその場ですぐに HTML 形式のレポートとなる。

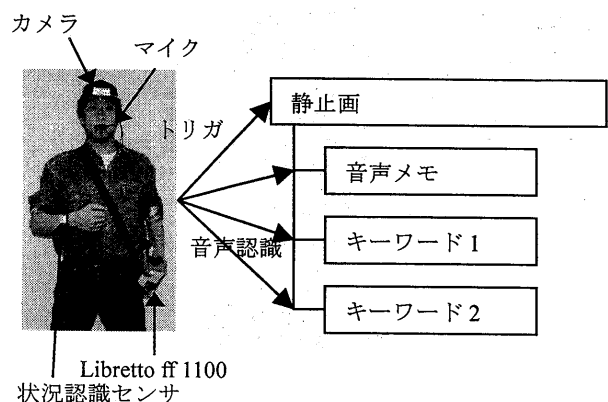


図1 システム概要

4. 実験

本システムの有効性を検証するために、1000人規模の展示会で本システムを展示し、ブースを訪れた方々のデータを実際に入力して即座に統計を作る実験を行った。

まず、来訪者の画像を2枚撮影し、その画像に対する音声メモ（3秒間）と、音声認識による音声キーワードを2種類入力した。キーワードには、来訪者の所属と、興味のある分野を入力した。ここで認識するキーワードは予めキーワード辞書に登録（10数個）しておいたものである。尚、これらの入力の切り替えは、前述の状況認識センサによりユーザの手首の動作を認識して行ったため、両手をふさぐことなく入力することができた。

このようにして入力された一連のマルチメディアデータはミニノートPCで関連付け、構造化され、HTMLファイル形式のデータとして生成されて、WebサーバへワイヤレスLANを用いてその場で送信される。送信された個人データは、Webブラウザ上で例えば図2のように表示される。ここで画像をクリックすると、その画像が撮影されたときに入力された音声メモが再生される。また音声認識により認識された二つのキーワードが画像の横にテキストで表示されている。このキーワードはデータの検索・統計等に利用することができる。今回の実験では、一日で約100名のデータを入力した。

本システムではさらに、音声キーワードを用いることにより、集められたマルチメディアデータから図3のようなサムネイル付きの統計グラフを生成して表示することもできる。左側のフレームの各メニューを選択することにより、所属別、興味別の一覧リストや、他の統計グラフも即座に参照することができる。このように、本システムを用いることによ

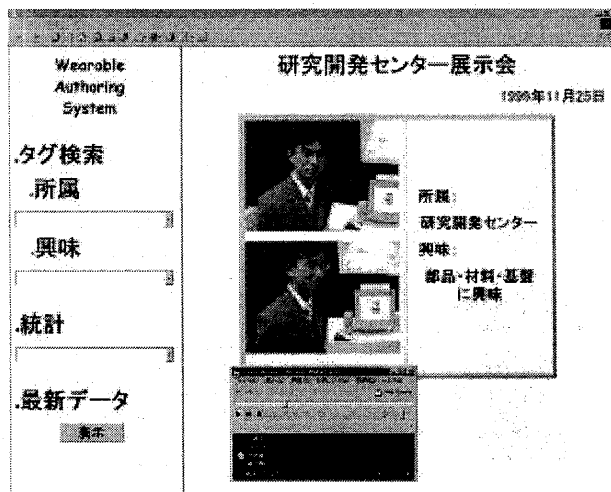


図2 保存された個人データの例

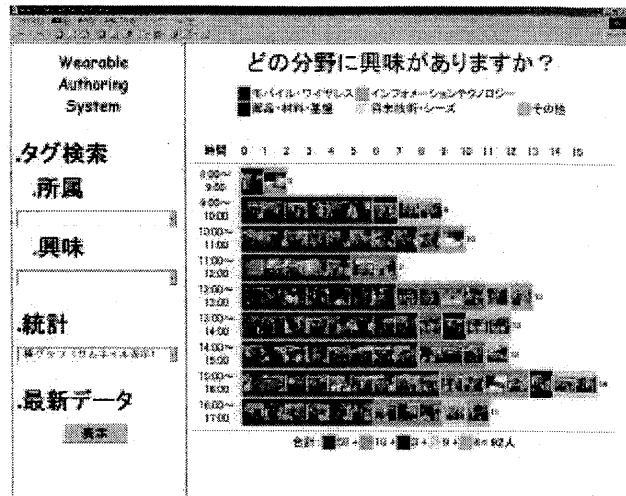


図3 統計グラフの一例
(サムネイルつき棒グラフ)

り、収集したデータを編集しなくても実用的なデータとして自動的に生成することができる。

図3の例では、興味のある分野ごとに色別で棒グラフ形式で表示している。縦軸は時間帯を示し、横軸はその時間帯に訪れた人の人数を示している。グラフ中の各サムネイルは、それぞれの個人データへリンクされているので、クリックするだけで、図2のような個人データを表示させることができる。

5. まとめ

本稿では、音声認識による音声キーワードを用いてマルチメディアデータを構造化することにより、編集することなくレポートやWebページを作成することができる、ウェアラブル・オーサリングシステムを提案した。今後は、音声認識に加えて、画像データ中の文字を文字認識させることにより、より有用なインデックスを利用することを計画している。また、獲得したデータを利用する際にも状況認識センサを利用し、ユーザの状況に応じて適切な情報を提示する方法についても検討を行っていく。

参考文献

- [1] B. Schilit, et al., "Context-Aware Computing Applications," IEEE Workshop on Mobile Computing Systems and Applications, 85-90, 1994.
- [2] S. Feiner, et al., "A Touring Machine: Prototyping 3D Mobile Augmented Reality Systems for Exploring the Urban Environment," ISWC'97, October 1997.
- [3] J. Pascoe, "Adding Generic Contextual Capabilities to Wearable Computers," ISWC'99, October 1999.
- [4] Y. Ohmori, "A Tool for Programmable Access to Organized Multimedia Data," IEEE APCHI'98, pp.249-252, 1998