サイバーゴーグル：画像情報からリアルタイムに実世界記述・検索を行うシステム

中山 英樹† 原田 達也† 國吉 康夫†
†東京大学大学院情報理工学系研究科

1 はじめに

近年，人間が日々の生活の中で観測する実世界情報をライフログとして常時記録し，その解析や整理を自動的に行う試みが盛んに行われている [1]．これは，人間の行動を支える高度に知的な情報処理を機械に実現させることに他ならず，デジタルアルバムの整理・検索，高齢者の記憶支援など幅広い産業応用上の成果を得ることが期待される．

そこで本研究では，視覚情報から環境中の事物を柔軟に認識・検索を行うサイバーゴーグルを提案する．これは，カメラを備えたゴーグルとタブレット型PCからなるシステムである．全体図を図1に示す．システムは，装着者の視界と同期したカメラ画像を取得し，提案する画像認識手法によりリアルタイムに物体認識を行い，認識結果をHMD (Head Mount Display) に出力する．同時に，認識結果と取得画像を合わせて視覚ログに随時保存する．また，検索の要求時には，視覚ログから該当する画像を選び，HMDへ出力する．

同様のシステムとしては [2] が挙げられるが，大規模な確率モデルの学習を必要とするため，認識対象数は限られている．より汎用的に使えるシステムの実現には，制約のない実世界画像から多様な物体やシーンを認識・検索できる技術が必要となる．これは，画像アノテーション・リトリーバル [3, 4] と呼ばれるが，既存手法は非常に膨大な計算処理を必要とするものばかりであり，実世界処理への応用は不可能であった．

そこで本システムでは，我々が開発した高速・高性能な画像認識・検索手法 [5] を実装する．我々の手法は，Corel5K [3] と呼ばれる画像データベースを用いた定量的な比較実験により，既存手法を精度・速度の両面で圧倒的に上回ることが示されており，前述のような実世界応用を初めて可能にするものである．

2 提案する画像認識・検索手法の概要

手法の目的は，画像と単語の関連性を学習することである．画像の特徴を $\mathbf{x}$，単語群の特徴を $\mathbf{w}$ とする．これらの同時確率 $P(\mathbf{x}, \mathbf{w})$ を，正準相関分析 (CCA) により学習する．$\mathbf{x}$,$\mathbf{w}$ についてCCAを行い，相関が最大となる新変量 (正準変量) $\mathbf{s} = A^T\mathbf{x}$，$\mathbf{t} = B^T\mathbf{w}$ を得る．これらは，$\mathbf{x}$ と $\mathbf{w}$ の関係において重要な特徴をとらえた圧縮表現となる．学習サンプルの正準変量を $\{\mathbf{s}_i\}_{i=1}^N$ とすると，$P(\mathbf{x}, \mathbf{w})$ は近似的に，

$$P(\mathbf{x}, \mathbf{w}) = \frac{1}{N}\sum P(\mathbf{x}|\mathbf{s}_i)P(\mathbf{w}|\mathbf{s}_i), \quad (1)$$

と表せる．この式の直感的な意味合いは，未知画像入力 $\mathbf{x}$ に対し，すべての学習サンプルとの類似度を計算し，この類似度で各サンプルについた単語群を重み付けしたものを答として出力すると解釈できる．$P(\mathbf{x}|\mathbf{s}_i)$ は，正準空間におけるサンプル間距離を用いて計算し，$P(\mathbf{w}|\mathbf{s}_i)$ はMBRM [3] などの言語モデルを用いる．

画像特徴としては，Color-HLAC特徴 [5] を用いる．単語群の特徴は，各単語が存在するか否かの1-0情報をベクトル状に並べたものを用いる．

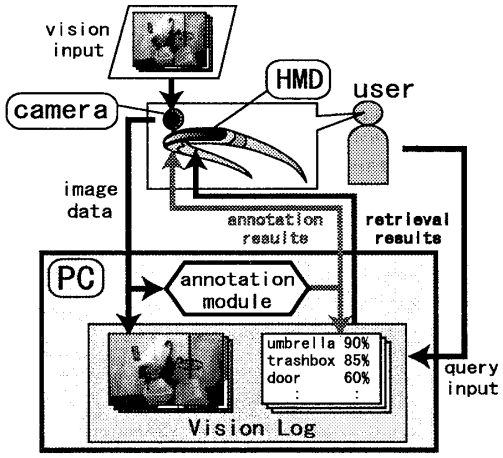

図1: Overall view of the cyber goggle system

3 サイバーゴーグルシステム

3.1 システム実装

本システムの実装図を図2に示す．このように，ゴーグルにカメラとHMDを搭載し，これらをタブレット型PC (CPU:Core2Solo 1.2GHz，メモリ1GB) に接続したものである．全ての処理はこのPCによって行われ，他のいかなる外部資源も必要としない．PCは非常に小

Cyber Goggle: Realtime description and retrieval of the real world using visual information
†H. NAKAYAMA, †T. HARADA, and †Y. KUNIYOSHI
†The University of Tokyo

型であるため，容易に持ち歩くことが可能であり，装着者の負担は小さい．

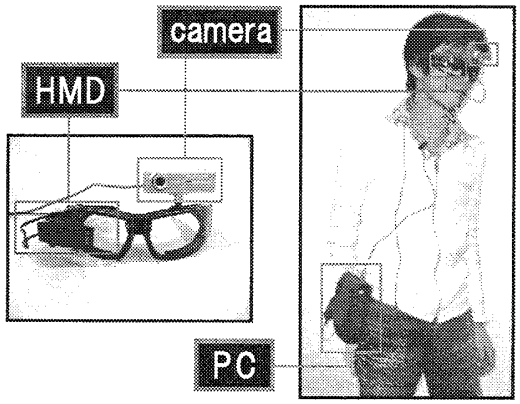

図 2: Implementation of the cyber goggle system

3.2 実験

ここでは，提案システムによる実世界中の事物の認識・検索が実際に可能であるかを検証する．図 3 に示すように，一般的な生活環境を想定した部屋をセットアップした．desk，chair のような家具にあたる大きな物体や，机の上に配置した books，clock などのような小物など，多様な物体を配置した．また，テーブル上の観葉植物については種類まで詳しく学習させる．このように，環境中におかれたさまざまな事物の認識・検索を行う．現在，用意している単語は 33 個である (図 3 には，代表的な物体のみ示してある)．

まず，あらかじめゴーグルを用いて大量の学習用画像を撮影しておき，それぞれについて数語のラベルを付け，学習サンプルとする．学習サンプル数は多いほどシステムの認識精度が向上するが，今回は約 2500 サンプルにより学習を行った．

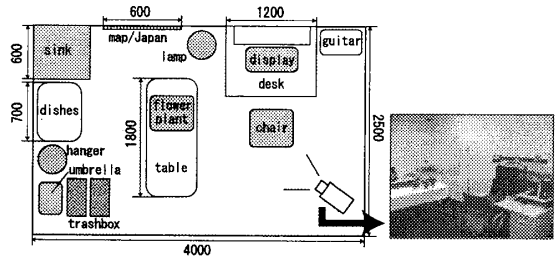

図 3: Layout of the experiment room

次に，学習した結果を用いて，実際にゴーグルを装着し，環境内の物体のリアルタイム認識を行った．視覚ログの保存は毎秒ごとに行い，画像自身と各単語の事後確率を記録する．認識結果の一例を図 4 に示す．単語の横の数値が，各単語の事後確率を表す．このように，実世界中の多様な物体について，柔軟に認識が可能であることが示された．

さらに，記録した視覚ログの中から，特定の物体が映った画像を検索させる実験を行った．この結果の一例を図 5 に示す．このように，さまざまな画像が含まれる視覚ログの中から所望の画像を正確に取り出せていることが分かる．

4 まとめ

本研究では，身の回りの物体の瞬間的な認識・検索を行うサイバーゴーグルシステムを提案した．これを可能としたものは，我々が提案した新しい画像認識・検索手法である．本手法は，精度・速度の両面で既存手法を圧倒しているため，限定的な計算資源においても高精度かつリアルタイムに認識・検索動作を行うことが可能となった．実際に提案手法をタブレット PC へ実装したサイバーゴーグルシステムの開発・検証により，その実効性を確認した．

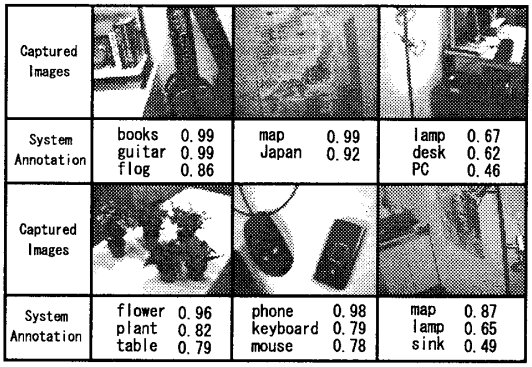

図 4: Example of the annotation

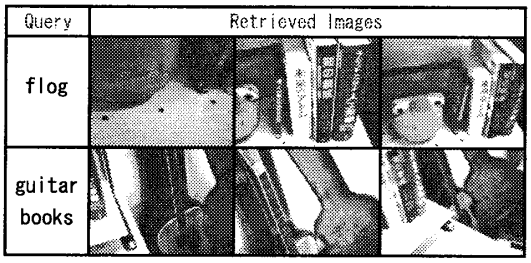

図 5: Example of the retrieval

参考文献

[1] K. Aizawa, Capture and retrieval of life log, ISWC Workshop on Ubiquitous Experience Media, pp.9-10, 2005.

[2] K. P. Murphy, A. Torralba, D. Eaton and W. T. Freeman, Sciliy workshop on object recognition, 2005.

[3] S. Feng, R. Manmatha and V. Lavrenko, Multiple bernoulli relevance models for image and video annotation, *Proc. CVPR*, pp.1063-69, 2004.

[4] G. Carneiro, A. B. Chan, P. J. Moreno and N. Vasconcelos, Supervised learning of semantic classes for image annotation and retrieval, *IEEE Trans. PAMI*, Vol.29, No.3, 2007.

[5] 中山 英樹, 原田 達也, 國吉 康夫, 画像・単語間概念対応の確率構造学習を利用した超高速画像認識・検索方法, 信学技報, Vol.107, No.384, PRMU2007-147, pp.65-70, 2007.