

# アンデス民族学画像を用いた類似画像検索システムの実装

2006MI122 西尾 翼

指導教員 河野 浩之

## 1 はじめに

南山大学人類学博物館の Web サイトに資料を閲覧・検索する環境として 2006 年, 草刈, 大原らの先行研究において, CMS ( Contents Management System ) とフォトアルバムモジュールによる博物館 Web サイトの構築が行われた。さらに 2008 年, 市川と河合らによってアーケイブシステムを備えた博物館 CMS の実装が行われた。これらのようなデジタルアーカイブの実施率は増加傾向にある。デジタルアーカイブ推進協議会 (JDAA) の調査によると 1997 年の 15.7 % に対して 2003 年には 29.4 % と着実に伸びている [1]。

本研究では DIGITAL MUSEUM の画像に対し類似画像検索を行う。南山大学人類学博物館のアンデス民族学画像は 4 万点以上の画像があるが, このような多くの画像から似た画像を検索するのに, 今まで通り一つずつ検索しては多くの時間が掛かる。その為, この南山大学人類学博物館の画像に対し類似画像検索を行い, 効率良く類似した画像を検索することが本研究の目的である。

## 2 類似画像検索システムの動向

一般に画像検索は, テキストに基づく画像検索 TBIR (Text-Based Image Retrieval) と, 画像の内容に基づく画像検索 CBIR (Content-Based Image Retrieval) に分けられる。TBIR はキーワードをメタデータとし, CBIR は画像の色, テクスチャ, 構図等をメタデータとして検索を行う [2]。

現在様々な類似画像検索システムが存在する。ここでは類似画像検索の先行研究として, Google Similar Image, GazoPa, QBIC, VisualSEEK, MARS, WebSEEK, 電子美術館の 7 つについて少しここで示す。表 1 は, 7 つの類似画像検索システムにおける検索方法と特徴量の分析方法を比較した結果である。

## 3 博物館としての CMS

本研究では汎用的な CMS を選択するため Joomla!, XOOPS, Drupal, Plone, Geeklog の五つの汎用 CMS に対して比較を行う。表 2 はこれら 5 つの CMS の機能である各々の言語, データベース, Gallery2 との連携性, 拡張機能の 4 項目の比較結果である [3]。

表 1 類似画像検索システム

検索システム	検索方法	分析手法
QBIC	CBIR	Wavelet 変換
Google Similar Image	TBIR	不明
GazoPa	CBIR	Wavelet 変換
WebSEEK	CBIR TBIR	自己組織化マップ Wavelet 変換
VisualSEEK	CBIR	Wavelet 変換
電子美術館	CBIR TBIR	正準相関分析
MARS	CBIR	relevance feedback

表 2 汎用 CMS の特徴・機能

	言語	DB	Gallery2	拡張
Joomla	PHP	MySQL		少
XOOPS	PHP	MySQL		多
Plone	Python	MySQL	不明	少
Drupal	PHP	MySQL PostgreSQL		多
Geeklog	PHP	MySQL		少

## 4 類似画像検索の実装

### 4.1 システムの構造

本研究は, CMS の Drupal を使用し, その上に Gallery2 と類似画像検索システムを導入する。そして類似画像検索システムと Gallery2 画像を連動することで, Gallery2 上の画像に対し類似画像検索をすることができる環境を実現した。システムの構造は図 1 のように構成した [4][5]。図 1 は Drupal 上の Gallery2 で表示した各画像からのリンクで PHP ファイルに移動し, そこでリンク元の画像名に対する類似画像を, 類似度順に表示するものである。類似度は, 画像に対する他画像の類似度を予め MySQL に格納しておく。そして表示の段階

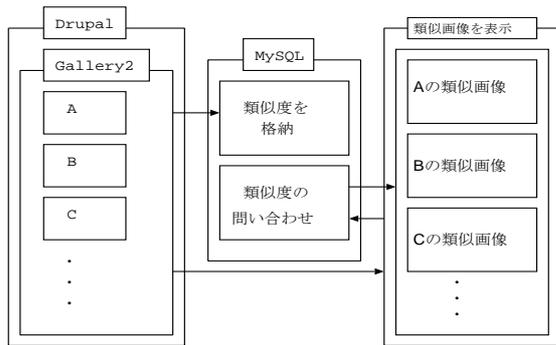


図 1 システムアーキテクチャ



図 2 類似画像検索結果の様子

で類似度を MySQL から取得し表示する。これにより類似画像検索を行う度に類似度を取得場合に比べ、大幅に表示までの時間を短縮した。また、リンク先に移った時は 2 秒以内に表示が完了するように処理量を抑えた。

#### 4.2 実装環境

本システムの実装環境は以下のような環境の元で行う。

- ・ PC のスペック
  - OS: Ubuntu 9.10
- ・ メモリ: 4GB
- ・ CPU のスペック
  - Inrel(R) Core(TM)2 Quad CPU
  - Q6600 @ 2.40GHz , 2.39GHz
- ・ 使用するソフト
  - Apache 2.0.50
  - PHP 5.0.1
  - MySQL 4.0.20
  - Drupal 5.20

#### 4.3 類似画像の表示

類似画像の表示は Gallery2 の各画像から PHP のページへ進み類似画像を表示する。類似度は OpenCV (Open Computer Vision) のサンプルコードの Sobel 法を用いた形状抽出を採用した [5]。

実装は画像名を PHP のページ側で受取り、その画像名に対する他画像名を類似度順に MySQL から取り出して表示する。また、類似度格納を含めたこれらのリンクは、Gallery2 のテンプレートを書き換えることでリンクを貼り、図 2 のように類似画像検索することができた。

実装により類似画像検索システムとして動作するが、類似度の取り込みにかかる時間が 475 点の画像に対しては約 10 時間かかった。類似性については博物館職員が類似画像と判断するもの画像を使い、植物、動物、人間、建物、風景の分類に分けて 10 回づつ実行した。結果、475 点では形状のみが類似していて画像として似ていな

いものが多く出たが、似ている画像も 1 割は確認された。

## 5 まとめ

システムの実装から、形状の自体の類似画像検索はできた。しかし、人の目で見えた類似画像は 1 割程度だった。今後、明確な形状を得ることができればより類似画像が検索できるだろう。また、本研究の目的である南山大学人類学博物館の画像資料の有意義な利用にもつながるであろう。

## 参考文献

- [1] 笠羽 晴夫, “デジタルアーカイブの構築と運用,” (株)水曜社, pp.22-23, 2004.
- [2] 佐々木 秀康, 清木 康, “画像データを対象とした特徴量類似度計量系によるメタデータ自動生成法の拡張,” 慶応義塾大学政策・メディア研究科, 慶応義塾大学環境情報学部, 社団法人電子情報通信学会, pp.159-160, 2003.
- [3] 田中裕也, 井ノ上憲司, 根本淳子, 鈴木克明, “オープンソース CMS の実証的比較分析と選択支援サイトの構築,” 日本教育工学会論文誌, pp.1-5, 2006.  
<http://www.geocities.jp/yuya0200005/files/ocet2.pdf> (accessed 2009/9/7).
- [4] M. V. Sudhamani, and C. R. Venugopal, “Grouping and Indexing Color Features for Efficient Image Retrieval,” pp.1-6, 2007.  
<http://www.waset.org/journals/ijamcs/v4/v4-3-25.pdf> (accessed 2009/9/7).
- [5] 斎藤雅紀, 中森伸行, “Wavelet 変換と Sobel フィルタを用いた歯科用 3 次元 CT 画像におけるノイズ低減法の開発画像圧縮回路,” 医用画像情報学会雑誌 Vol.24, No.1, pp.1-5, 2007.