A Multimedia Data Base Browsing System

Massimiliano Albanese
Carmine Cesarano
Antonio Picariello

Dipartimento di Informatica e Sistemistica
University of Naples “Federico II”
Naples, Italy

Introduction

- Browsing of large multimedia data bases
  - Availability of great amounts of data
  - Complexity of the retrieval techniques
- Browsing and retrieval from multimedia data bases is usually based on the similarity of
  - low-level features
    - e.g. color and texture in the image realm, motion information in the video domain
  - high-level features
    - semantic descriptors or textual annotations

Motivating example

- We consider the case of a “virtual museum”
  - A “virtual museum” is a museum that offers a web based access to a collection of digital reproduction of paintings
- Our approach can be extended to any kind of multimedia object



Contribution

- A Multimedia Data Base Browsing System that gives useful recommendations to the user based on
  - image analysis and features extraction
  - human-created annotations and taxonomic knowledge bases
  - user preferences, represented by the navigational behavior of current and past users

System Architecture

- Usage Log
  - stores log data of users' browsing sessions
- Pattern Discovery Subsystem
  - recognizes the users having the most similar behavior to the current users
- Recommendation Subsystem
  - produces a ranked list of recommended items (paintings)

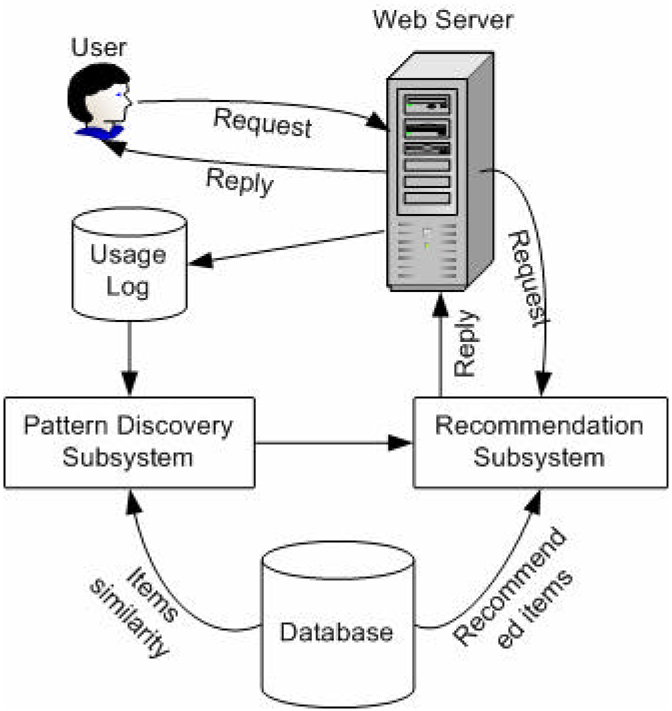


Image Comparison

- Wavelet Transform
  - reduces the amount of data to be analyzed
  - provides a suitable color and texture representation
- Color Distance
  - comparison of color histograms
    - computed on the low-pass components (a smoothed copy of the original picture)

$$d_{col}(I_1, I_2) = 1 - \sum_{b=1}^{B} \min\left(h_b^{I_1}, h_b^{I_2}\right) / \sum_{b=1}^{B} h_b^{I_1}$$

Image Comparison

- Texture features

$$\Sigma_C^2 = \{\sigma_{X,Y}^2\} \qquad \sigma_{X,Y}^2 = \sum_{x,y} \left\{ \frac{1}{|D_p|/4} \sum_{k=1}^{3} X_k(x,y) Y_k(x,y) \right\}$$

- Texture Distance

$$d_{tex}(I_1, I_2) = \frac{1}{R} \sum_{i=1}^{|\Sigma^2|} \frac{\left| \Sigma_{C_1}^2[i] - \Sigma_{C_2}^2[i] \right|}{\min\left(\left|\Sigma_{C_1}^2[i]\right|, \left|\Sigma_{C_2}^2[i]\right| \right)}$$

  - R is a normalization factor
  - $|\Sigma^2|$ is the number of features

Features Based Similarity

- The Features Based Distance and Similarity between two images I_1 and I_2 are defined as

$$d_{features}(I_1, I_2) = \alpha_{col} \cdot d_{col}(I_1, I_2) + \alpha_{tex} \cdot d_{tex}(I_1, I_2)$$

$$S_F(I_1, I_2) = 1 - d_{features}(I_1, I_2)$$

- α_{col} and α_{tex} are two weighting factors
  - chosen by maximizing the correlation between human judged similarity and S_F
    - $\alpha_{col} = 0.67$ and $\alpha_{tex} = 0.33$

Taxonomies

- A taxonomy is a hierarchical concept network
- Given an application specific taxonomy T, we define an Object Semantic Description as

  $OSD = (TA, NTA)$
- $TA = (A_1, \ldots, A_\tau)$ is an ordered tuple of attributes that assumes values represented by nodes of T
- $NTA = (A_1^*, \ldots, A^*_{\tau^*})$ is an ordered tuple of attributes that assumes values not represented by nodes of T
- Let O denote the set of all the objects in the collection, defined as triples $(OID, OSD, PhyObj)$

Taxonomy Based Similarity

- The Features Based Similarity and Distance between two objects o_1 and o_2 are defined as

$$S_T(o_i, o_j) = \frac{1}{\tau} \cdot \sum_{k=1}^{\tau} e^{-\alpha \cdot l(a_k^i, a_k^j)} \cdot \left(1 - e^{-\beta \cdot d(a_k^i, a_k^j)}\right)$$

$$d_{taxonomy}(o_i, o_j) = 1 - S_T(o_i, o_j)$$

- a_k^i and a_k^j are the values of A_k for o_i and o_j
- $l(a_k^i, a_k^j)$ is the shortest path length and $d(a_k^i, a_k^j)$ the depth of the subsumer between a_k^i and a_k^j
- α and β are parameters scaling the contribution of shortest path length and depth respectively

Features and taxonomy
A combined metric

- The combined distance metric used to index the images into an M-tree structure and the corresponding similarity measure are defined as

$$d_M(I_i, I_j) = \alpha_F \cdot d_{features}(I_i, I_j) + \alpha_T \cdot d_{taxonomy}(I_i, I_j)$$

$$S_M(I_i, I_j) = 1 - d_M(I_i, I_j)$$

- α_F and α_T are two weighting factors
  - chosen by maximizing the correlation between human judged similarity and S_F
    - $\alpha_F = 0.62$ and $\alpha_T = 0.38$

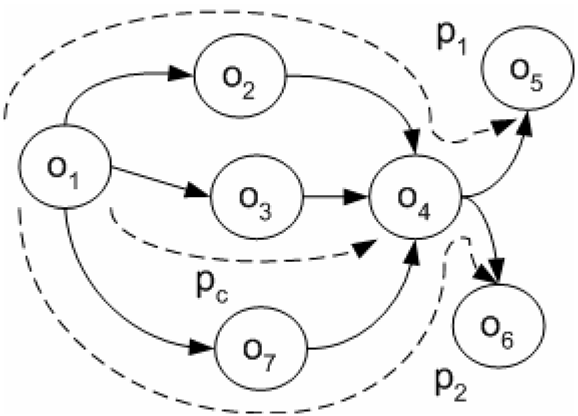
User preferences
Usage patterns

- Recommendations for current users are based on the behavior of past users
- Behavior is represented through the usage patterns of the users

- A Usage pattern p of length k is defined as

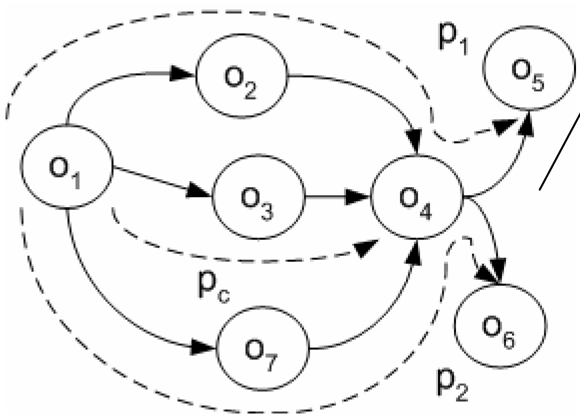
$$p = (o_{i_1}, o_{i_2}, \ldots, o_{i_k}),\ with\ o_{i_j} \in \mathcal{O}\ \forall j \in [1, k]$$

- Let *P* denote the set of all the usage patterns of past visitors



Comparison of usage patterns

- Some distances (e.g. Levenshtein) have been defined to evaluate the distance between sequences of symbols from a given alphabet Σ
  - Only the alignment of the symbols is taken into account



The Levenshtein distance between patterns $p_1=(o_1,o_3,o_4,o_5)$ and $p_2=(o_1,o_2,o_4,o_5)$ is equal to 1, whatever o_3 and o_2 are

- Our approach
  - Evaluate the similarity between patterns based on the similarity between objects

Local similarity of patterns
The proposed algorithm

```
function local-similarity(p1,p2)
    p1 and p2 are two patterns of length m and n respectively
    D is a two-dimensional array with m + 1 rows and n + 1 columns
begin
    for j ← 0 to n do
        D[0, j] ← 0
    end for
    for i ← 0 to m -1 do
        D[i + 1, 0] ← 0
        for j ← 0 to n do
            D[i + 1, j + 1] ← max { 0, D[i, j] + Sub(p1[i], p2[j]),
                                    D[i, j+1] + Del(p1[i], p2[j]), D[i + 1, j] + Ins(p2[j], p1[i]) }
        end for
    end for
    return max_{i, j}{D[i, j]} / min{m, n}
end
```

Sub(), *Ins*() and *Del*() functions

- *Sub*(), *Ins*() and *Del*() are the functions used to reward/penalize an alignment

$$Sub(p_1[i], p_2[j]) = \frac{S_M(o_{k_i}, o_{l_j}) - \delta}{1 - \delta}$$

$$Ins(p_2[j], p_1[i]) = \frac{\min\{S_M(o_{k_i}, o_{l_j}), S_M(o_{k_{i+1}}, o_{l_j})\} - 1}{(1 - \delta)/\delta}$$

$$Del(p_1[i], p_2[j]) = Ins(p_1[i], p_2[j])$$

- δ is a threshold defined as

$$\delta = (\lg|\mathcal{O}| - 0.4)/\lg|\mathcal{O}|$$

Recommendation process
Selection of candidate objects

- Consider the following sets

$$\mathcal{P}_\gamma = \{p \in \mathcal{P} \mid \text{local-similarity}(p, p_c) \geq \gamma\}$$

$$\mathcal{O}_\gamma = \{o \in \mathcal{O} \mid \exists p \in \mathcal{P}_\gamma, \text{next}_p(p_c) = o\}$$

- The set of candidate objects to be recommended is defined as

$$\mathcal{O}_c = \mathcal{O}_\gamma \cup \mathsf{NN}(o_c, k)$$

- γ is a threshold defined as $\gamma = (|\mathcal{P}| - 0.2)/|\mathcal{P}|$
- NN(o_c, k) is a k nearest neighbors query

Recommendation process
Pattern Based Similarity

- Consider the following sets

$$\mathcal{P}_i = \{p \in \mathcal{P}_\gamma \mid \text{next}_p(p_c) = o_i\}, \forall o_i \in \mathcal{O}_c$$

- The Pattern Based Similarity of an object o_i w.r.t. to the current path p_c is defined as

$$S_P(o_i) = \frac{\sum_{p \in \mathcal{P}_i} \textit{local-similarity}(p, p_c)}{\max_i \left\{ \sum_{p \in \mathcal{P}_i} \textit{local-similarity}(p, p_c) \right\}}$$

- $S_P(o_i)$ can be thought as the likelihood that the user may be interested in o_i, given her current usage pattern and the usage patterns of past visitors

Recommendation process
Recommendation Grade

- The Recommendation Grade of an object o_i, given the current pattern p_c and the last element o_c in p_c, is defined as

$$\rho(o_i) = \alpha_M \cdot S_M(o_i, o_c) + \alpha_P \cdot S_P(o_i)$$

- α_M and α_P are two weighting factors, defined as functions of the length n_c of the current pattern

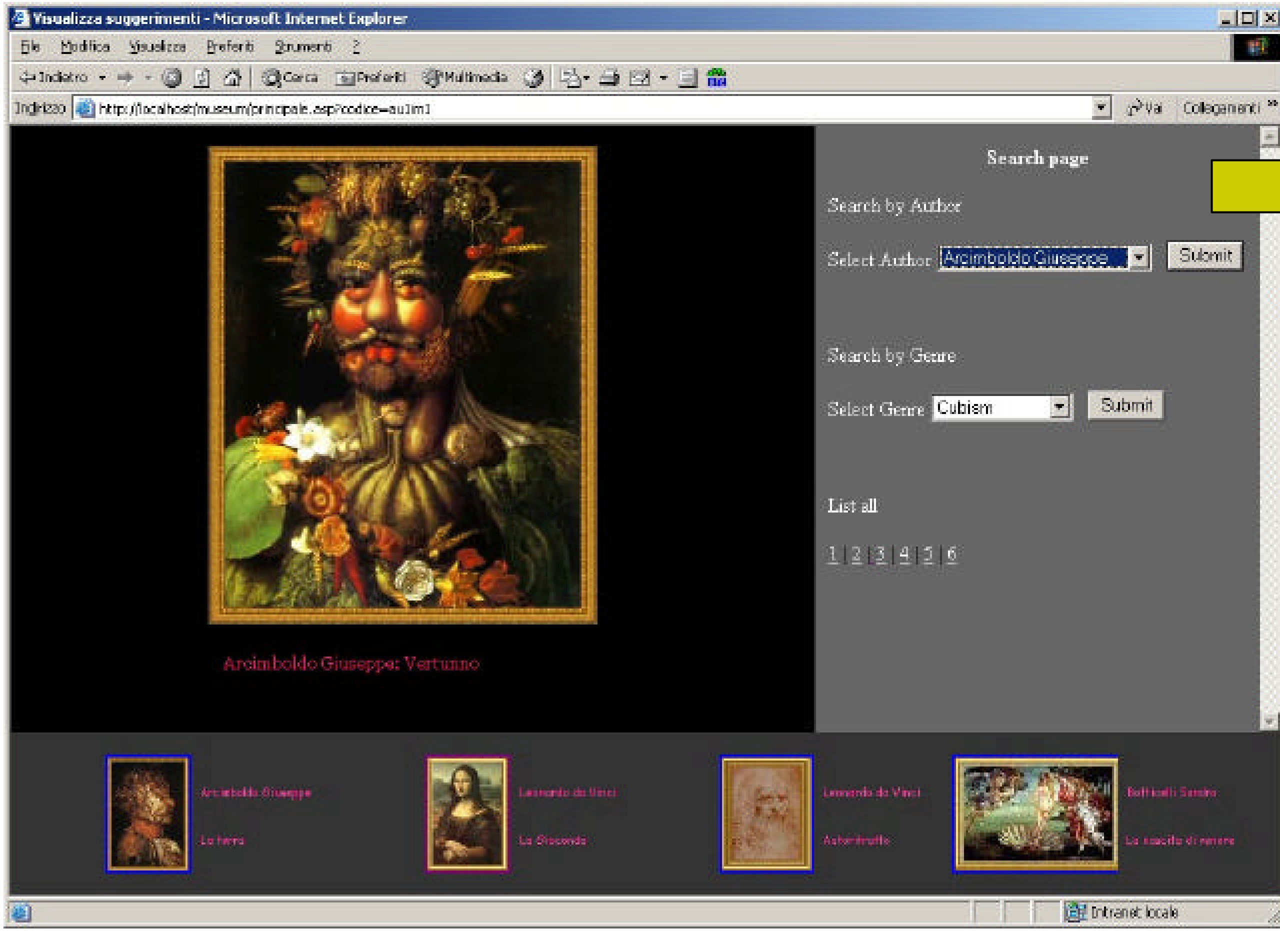
$$\alpha_M = 1/n_c \qquad \alpha_P = (n_c - 1)/n_c$$

- The k objects in O_c exhibiting the highest values of ρ are recommended to the user

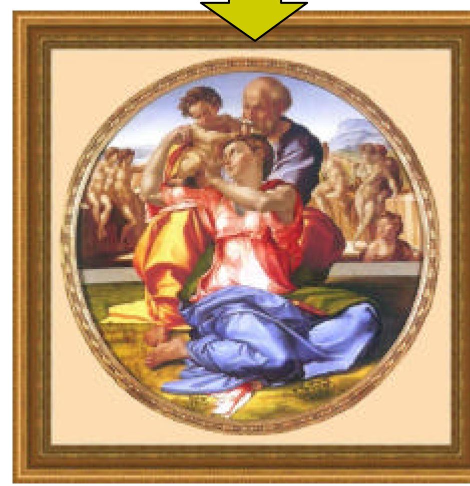
Scale issues

- Two main scale issues
  - Size of the image (object) collection
    - Addressed using an M-tree to index the images
    - The d_M distance has been adopted to partition the metric space
  - Size of the usage pattern collection
    - Addressed using an M-tree to index the patterns
    - The metric space has been partitioned using the distance $d = (1 - \text{local-similarity})$
    - The set P_γ can be determined as
      - range(p_c, $1 - \gamma$)

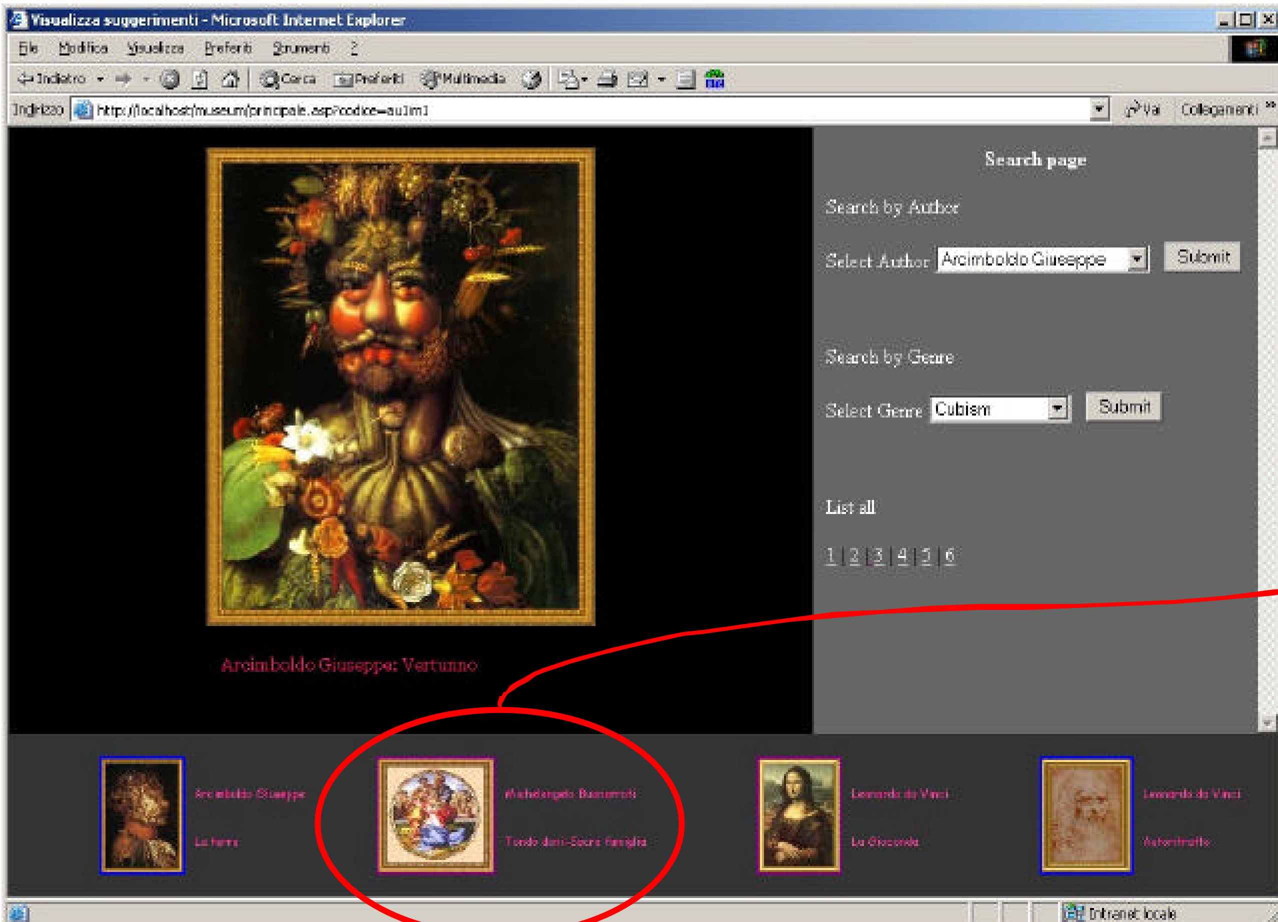
Experiments
Example of use



The user selects a picture other than the recommended ones



Experiments
Example of use



Next users with a similar behavior will be invited to see that picture

Experiments
Qualitative evaluation

- Step 1
  - We asked a group of about 60 people to use the system for some days, in order to collect a significant amount of usage patterns
- Step 2
  - We asked a different group of 20 people to browse the collection using
    - the standard search capabilities
    - the assistance of the recommender system
  - 73% of the people found the system helpful, while the remaining 27% did not appreciate significant differences

Conclusions

- We have shown that the proposed system provides the following interesting insights:
  - the recommendation algorithm does not use any preliminary knowledge about the users' behavior
  - the recommendations are produced using both visual and semantic descriptions
  - the impact on the users at this early stage of the experimentations is promising
- Open issues
  - Extending analysis and experiments to more general scenarios and different kind of multimedia data, such as video
  - How to create an adequate semantic taxonomy for different realms

Acknowledgements

- This work has been carried out partially under the financial support of the Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR) in the framework of the FIRB Project “Middleware for advanced services over large-scale, wired-wireless distributed systems (WEB-MINDS)”