Patent Image Retrieval based on Concept Extraction and Classification

Stefanos Vrochidis
Anastasia Mountzidou
Ioannis Kompatsiaris

Centre for Research and Technology Hellas
Informatics and Telematics Institute

IRF Symposium 2011
Vienna, 7 June 2011
Overview

• Introduction
• Concept Extraction Framework
• Dataset and Concepts
• Approach
• Demo
• Results
• Conclusions
• Future Work
Introduction

- Image Search is mostly based on:
  - visual low level features
  - textual annotations
  - combination of the above

- State of the art image retrieval systems attempt to extract high level concepts from images based on:
  - Machine learning using manually annotated training data and low level features
  - Combination and fusion of heterogeneous information

- The main problem faced is the “semantic gap” between low and high level features
Introduction

• Patent searchers are usually searching based on concept information

• Example [1]:

  Disclosure reads:
  A dancing shoe with a rotatable heel to allow rapid pivoting about your heel. In a preferred embodiment, the heel should have ball bearings.

  The gist:
  Concept 1: Dancing shoe
  Concept 2: Rotating heel
  Refined Concept 2: Rotating Heel with ball bearings

• It would be helpful for patent searchers to search based on concepts.

Concept Extraction Framework

- Supervised Machine learning based framework
- Trained with textual and visual low level features
- Requires Manually annotated Training Data
Dataset and Concepts

- Patent Images
  - A43B IPC sub-class
  - Contain “Parts of Footwear”

- Concepts*
  - Cleat
  - Ski boot
  - High heel
  - Lacing Closure
  - Spring Heel
  - Tongue

* The selection of concepts was done with the help of Dominic DeMarco
Concepts

• Cleat
  • Description: A short piece of rubber, metal etc attached to the bottom of a sports shoe used mainly for preventing someone from slipping
  • IPC subclass: A43B5/18S

• Ski boot
  • Description: A specially made boot that fastens onto a ski
  • IPC subclass: A43B5/04
Concepts

• **High Heel**
  • Description: Shoes with high heels
  • IPC subclass: A43B21

• **Lacing closure**
  • Description: A cord that is drawn through eyelets or around hooks in order to draw together the two edges of a shoe
  • IPC subclass: A43B5/04
Concepts

• **Spring Heel**
  • Description: Heels with metal springs
  • IPC subclass: A43B21/30

• **Tongue**
  • Description: The part of a shoe that lies on top of your foot, under the part where you tie it
  • IPC subclass: A43B23/26
Dataset Statistics

<table>
<thead>
<tr>
<th>Concept</th>
<th>All figures</th>
<th>Train Data</th>
<th>Test Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cleat</td>
<td>148</td>
<td>89</td>
<td>59</td>
</tr>
<tr>
<td>Ski boot</td>
<td>123</td>
<td>74</td>
<td>49</td>
</tr>
<tr>
<td>High heel</td>
<td>148</td>
<td>89</td>
<td>59</td>
</tr>
<tr>
<td>Lacing Closure</td>
<td>117</td>
<td>71</td>
<td>46</td>
</tr>
<tr>
<td>Spring Heel</td>
<td>106</td>
<td>64</td>
<td>42</td>
</tr>
<tr>
<td>Tongue</td>
<td>124</td>
<td>75</td>
<td>49</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>766</strong></td>
<td><strong>352</strong></td>
<td><strong>304</strong></td>
</tr>
</tbody>
</table>

- Creation of training/testing set
  - Testing/Training ratio = 3/5
  - Positive/Negative ratio = 1/3
Visual Features

- Extraction of Adaptive Hierarchical Density Histograms (AHDH) as visual feature vectors [2]
  - Global visual features based on the pixel distribution of a drawing
  - Low dimension feature vector (~100 features)

Textual Features

• Textual information extraction

Example: Patent US 20020152637 A1

FIG. 7 shows the reversible tongue containing a pocket in its upper half, and which may be secured by Velcro, or the like, into closure

• Bag of words technique
  • Indexing textual information using Lemur [3]
  • Stemming using Porter stemmer
  • Creation of Lexicon using training data
  • Feature vector includes lexicon term weights
    • <boot snowboard illustr tongu footwear heel...>
    • [0 0 0 0.0909091 0 0...]

Support Vector Machines

• Support Vector Machines (SVM) constitute a set of supervised learning methods used for:
  • classification
  • Regression

• When a set of training positive and negative examples is available, a SVM training algorithm builds a model that predicts in which category a new example falls into.
• SVM constructs a hyperplane in a high or infinite dimensional space.
• The best separation is achieved by the hyperplane that has the largest distance from the nearest training datapoints.
• We employed C-SVC SVM with a polynomial kernel [4]

Approach

- We trained one classifier for each concept
- The following cases are considered
  - Visual
    - SVMs were trained only with visual features (AHDH)
  - Visual extended
    - Extension of visual case
    - The output of all classifiers forms a vector and is passed to a new classifier to yield the final score
  - Textual
    - SVMs were trained only with textual features
  - Visual + Textual
    - SVMs were trained with a feature vector containing visual and textual features
    - 200 features (100 visual and 100 textual)
Approach

Feature Vector

Cleat Classifier → Cleat score → Cleat score Classifier
Ski boot Classifier → Ski boot score → Ski boot score Classifier
High Heel Classifier → High Heel score → High Heel score Classifier
Lacing Closure Classifier → Lacing Closure score → Lacing Closure score Classifier
Spring Heel Classifier → Spring Heel score → Spring Heel score Classifier
Tongue Classifier → Tongue score → Tongue score Classifier

Max score output

Visual
Textual
Textual + Visual

Visual extended

Max score output
Time for the demo!

http://mklab-services.iti.gr/patmediac
## Results-Cleat

<table>
<thead>
<tr>
<th>Features</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual</td>
<td>88,16%</td>
<td>81,08%</td>
<td>50,85%</td>
<td>62,5%</td>
</tr>
<tr>
<td>Textual</td>
<td>85,86%</td>
<td>67,39%</td>
<td>52,54%</td>
<td>59,05%</td>
</tr>
<tr>
<td>Visual ext.</td>
<td>74,34%</td>
<td>41,88%</td>
<td>83,05%</td>
<td>55,68%</td>
</tr>
<tr>
<td>Visual+Textual</td>
<td>92,11%</td>
<td>87,23%</td>
<td>69,49%</td>
<td>77,36%</td>
</tr>
</tbody>
</table>
Results-Ski boot

<table>
<thead>
<tr>
<th>Features</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual</td>
<td>93,09%</td>
<td>76,92%</td>
<td>81,63%</td>
<td>79,21%</td>
</tr>
<tr>
<td>Textual</td>
<td>85,86%</td>
<td>67,39%</td>
<td>52,54%</td>
<td>59,05%</td>
</tr>
<tr>
<td>Visual ext.</td>
<td>92,43%</td>
<td>79,55%</td>
<td>71,43%</td>
<td>75,27%</td>
</tr>
<tr>
<td>Visual+Textual</td>
<td>95,39%</td>
<td>77,78%</td>
<td>100%</td>
<td>87,5%</td>
</tr>
</tbody>
</table>

![Diagram of Ski Boot](image)
## Results-High Heel

<table>
<thead>
<tr>
<th>Features</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual</td>
<td>89.47%</td>
<td>68.49%</td>
<td>84.75%</td>
<td>75.76%</td>
</tr>
<tr>
<td>Textual</td>
<td>90.46%</td>
<td>70.83%</td>
<td>86.44%</td>
<td>77.86%</td>
</tr>
<tr>
<td>Visual ext.</td>
<td>85.86%</td>
<td>58.33%</td>
<td>94.92%</td>
<td>72.26%</td>
</tr>
<tr>
<td>Visual+Textual</td>
<td>90.79%</td>
<td>69.62%</td>
<td>93.22%</td>
<td>79.71%</td>
</tr>
</tbody>
</table>
# Results - Lacing Closure

<table>
<thead>
<tr>
<th>Features</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual</td>
<td>88.82%</td>
<td>87.5%</td>
<td>30.43%</td>
<td>45.16%</td>
</tr>
<tr>
<td>Textual</td>
<td>91.78%</td>
<td>68.09%</td>
<td>76.19%</td>
<td>71.91%</td>
</tr>
<tr>
<td>Visual ext.</td>
<td>91.45%</td>
<td>83.33%</td>
<td>54.35%</td>
<td>65.79%</td>
</tr>
<tr>
<td>Visual+Textual</td>
<td>89.8%</td>
<td>74.19%</td>
<td>50%</td>
<td>59.74%</td>
</tr>
</tbody>
</table>
Results-Heel with spring

<table>
<thead>
<tr>
<th>Features</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual</td>
<td>90,79%</td>
<td>79,17%</td>
<td>45,24%</td>
<td>57,58%</td>
</tr>
<tr>
<td>Textual</td>
<td>91,78%</td>
<td>68,09%</td>
<td>76,19%</td>
<td>71,91%</td>
</tr>
<tr>
<td>Visual ext.</td>
<td>89,8%</td>
<td>60,78%</td>
<td>73,81%</td>
<td>66,66%</td>
</tr>
<tr>
<td>Visual+Textual</td>
<td>94,74%</td>
<td>96,43%</td>
<td>64,29%</td>
<td>77,15%</td>
</tr>
</tbody>
</table>
## Results - Tongue

<table>
<thead>
<tr>
<th>Features</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual</td>
<td>63.16%</td>
<td>29.14%</td>
<td>89.8%</td>
<td>44%</td>
</tr>
<tr>
<td>Textual</td>
<td>96.05%</td>
<td>87.76%</td>
<td>87.76%</td>
<td>87.76%</td>
</tr>
<tr>
<td>Visual ext.</td>
<td>89.47%</td>
<td>69.77%</td>
<td>61.22%</td>
<td>65.22%</td>
</tr>
<tr>
<td>Visual + Textual</td>
<td>96.71%</td>
<td>88.24%</td>
<td>91.84%</td>
<td>90%</td>
</tr>
</tbody>
</table>
Results

- Average Results for all concepts

<table>
<thead>
<tr>
<th>Features</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual</td>
<td>85.58%</td>
<td>70.38%</td>
<td>63.78%</td>
<td>60.7%</td>
</tr>
<tr>
<td>Textual</td>
<td>90.3%</td>
<td>71.59%</td>
<td>71.94%</td>
<td>71.25%</td>
</tr>
<tr>
<td>Visual ext.</td>
<td>87.22%</td>
<td>65.6%</td>
<td>73.13%</td>
<td>66.81%</td>
</tr>
<tr>
<td>Visual + Textual</td>
<td>93.25%</td>
<td>82.24%</td>
<td>78.14%</td>
<td>78.58%</td>
</tr>
</tbody>
</table>
Conclusions

- Combination of visual and textual information performs better.
- Classification based solely on visual information is still very satisfactory.
- Visual extended approach reports an improved F-score compared to visual.
- Classification based on visual features could fail when two visually similar images are described with different concepts.
- SVM testing results are better than query by visual example results due to the training process.
- Training requires manual effort due to annotation and segmentation.
- Automatic segmentation could be supported, however an error (~20%) might be introduced.
- The concept retrieval module could be a part of a larger patent retrieval framework.
Future Work

• Produce results for more concepts.
• Realize the same framework for bigger datasets and for different IPC classes.
• Test performance in the case of automatically segmented images (i.e. of lower quality).
• Combine more efficiently text and visual information
  • give more weight to visual or textual description depending on the concept type
• Investigate late fusion techniques.
Feel free to test the demo!
http://mklab-services.iti.gr/patmediac

Thank you!
http://mklab.iti.gr