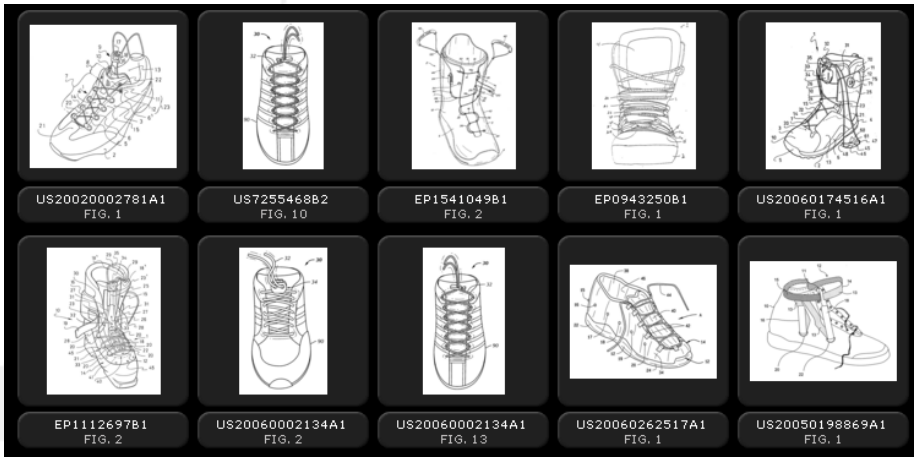
Patent Image Retrieval based on Concept Extraction and Classification



Stefanos Vrochidis
Anastasia Moumtzidou
Ioannis Kompatsiaris

Centre for Research and Technology Hellas

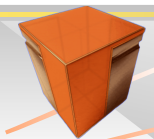
Informatics and Telematics Institute

IRF Symposium 2011

Vienna, 7 June 2011

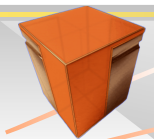
Overview

- 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

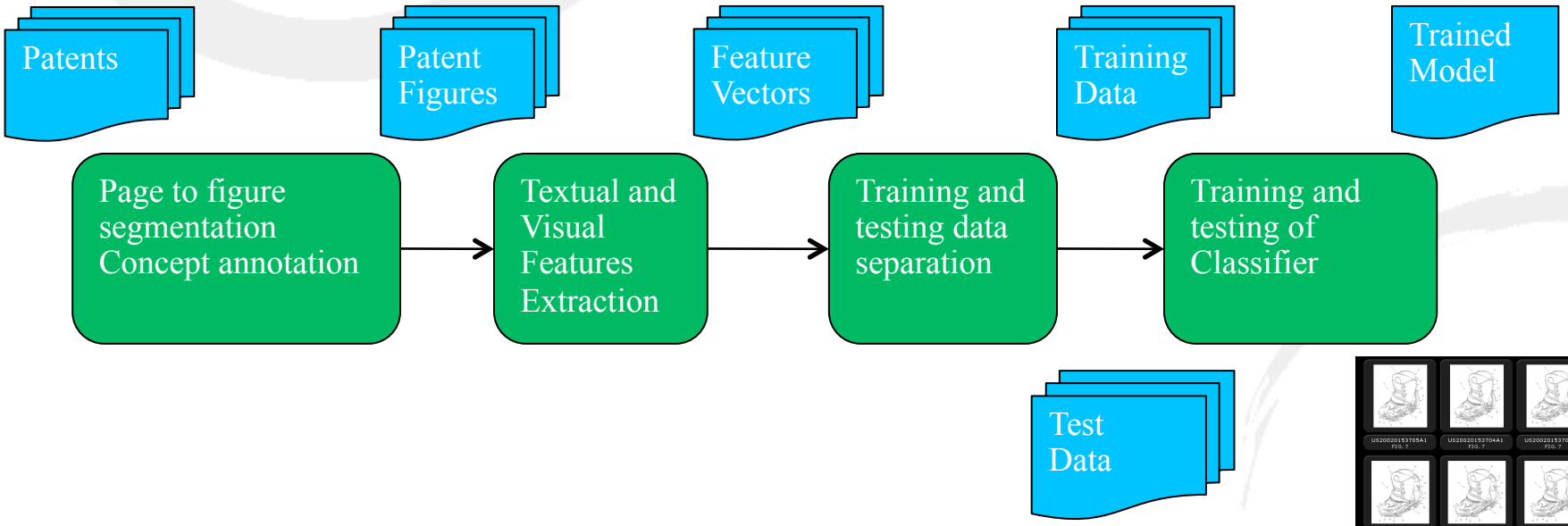
Refined Concept 2: Rotating Heel with ball bearings

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

[1] Dominic DeMarco, Mechanical Patent Searching: A Moving Target - PIUG 2010 Annual Conference

Concept Extraction Framework

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

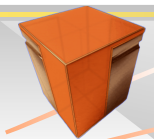


Dataset and Concepts

- Patent Images
  - A43B IPC sub-class
  - Contain "Parts of Footwear"

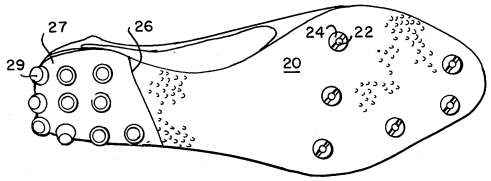
- Concepts*
  - Cleat
  - Ski boot
  - High heel
  - Lacing Closure
  - Spring Heel
  - Tongue

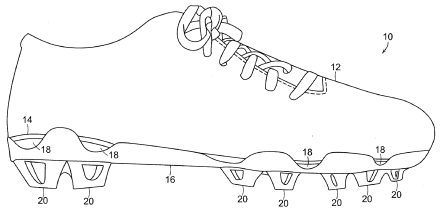
* The selection of concepts was done with the help of Dominic DeMarco



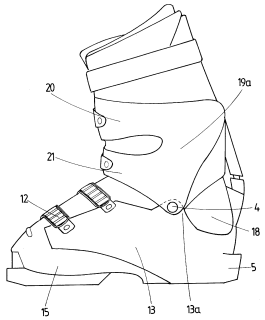
Concepts

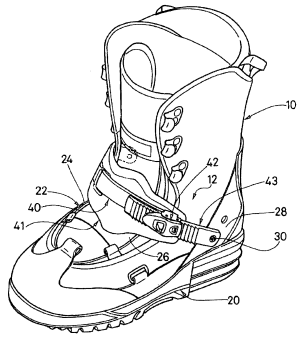
- Cleat
  - Description: A short piece of rubber, metal etc attached to the bottom of a sports shoe used mainly for preventing someone from slipping
  - IPC subclass: A43B5/18S





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

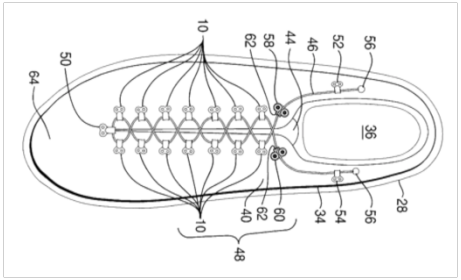


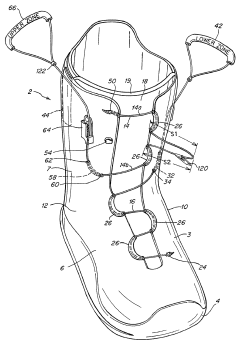


Concepts

- High Heel
  - Description: Shoes with high heels
  - IPC subclass: A43B21

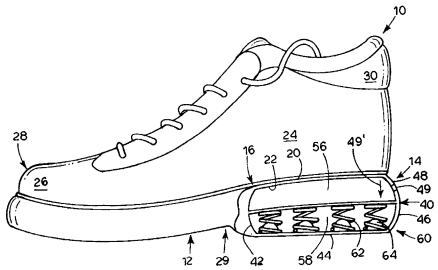
- Lacing closure
  - Description: A cord that is drawn through eyelets or around hooks in order to draw together the two edges of a shoe
  - IPC subclass: A43B5/04

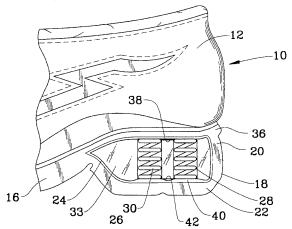




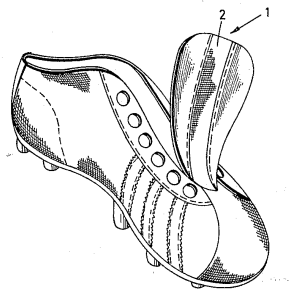
Concepts

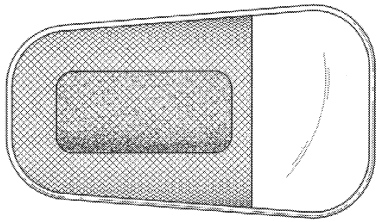
- Spring Heel
  - Description: Heels with metal springs
  - IPC subclass: A43B21/30





- 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

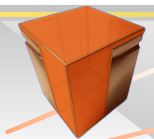




Dataset Statistics

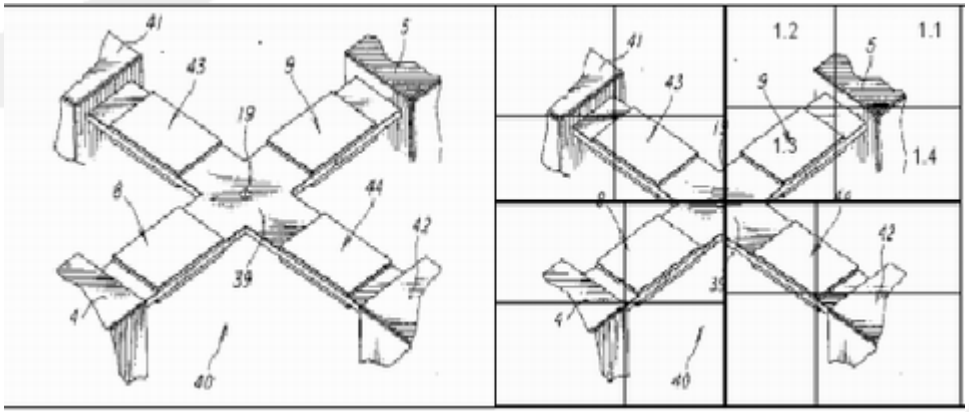
Concept	All figures	Train Data	Test Data
Cleat	148	89	59
Ski boot	123	74	49
High heel	148	89	59
Lacing Closure	117	71	46
Spring Heel	106	64	42
Tongue	124	75	49
Total	**766**	**352**	**304**

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



Visual Features

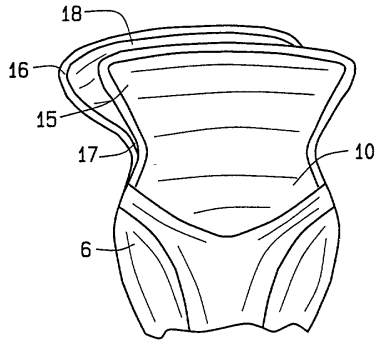
- Extraction of Adaptive Hierarchical Density Histograms (AHDH) as visual feature vectors [2]
  - Global visual features based on the pixel distribution of a drawing
  - Low dimension feature vector (~100 features)



[2] P. Sidiropoulos, S. Vrochidis, I. Kompatsiaris, "Content-Based Binary Image Retrieval using the Adaptive Hierarchical Density Histogram", Pattern Recognition Journal, Elsevier, Volume 44, Issue 4, pp 739-750, April 2011.

Textual Features

- Textual information extraction



Example: Patent US 20020152637 A1

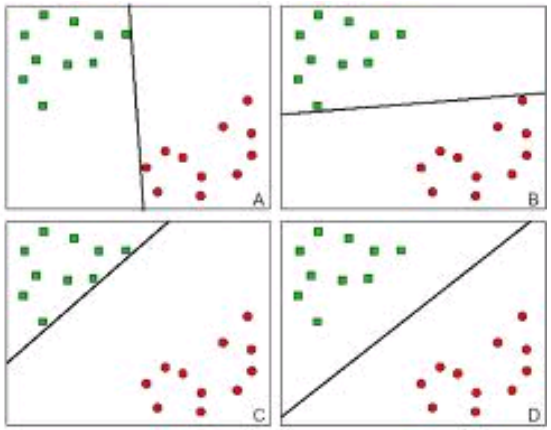
> FIG. 7 shows the reversible tongue containing a pocket in its upper half, and which may be secured by Velcro, or the like, into closure

- 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...]

[3] http://www.lemurproject.org/

Support Vector Machines

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

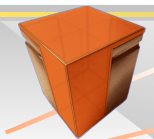


- When a set of training positive and negative examples is available, a SVM training algorithm builds a model that predicts in which category a new example falls into.
- SVM constructs a hyperplane in a high or infinite dimensional space.
- The best separation is achieved by the hyperplane that has the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

[4] LibSVM: http://www.csie.ntu.edu.tw/~cjlin/libsvm/

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
Ski boot Classifier
High Heel Classifier
Lacing Closure Classifier
Spring Heel Classifier
Tongue Classifier

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

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

Max score output

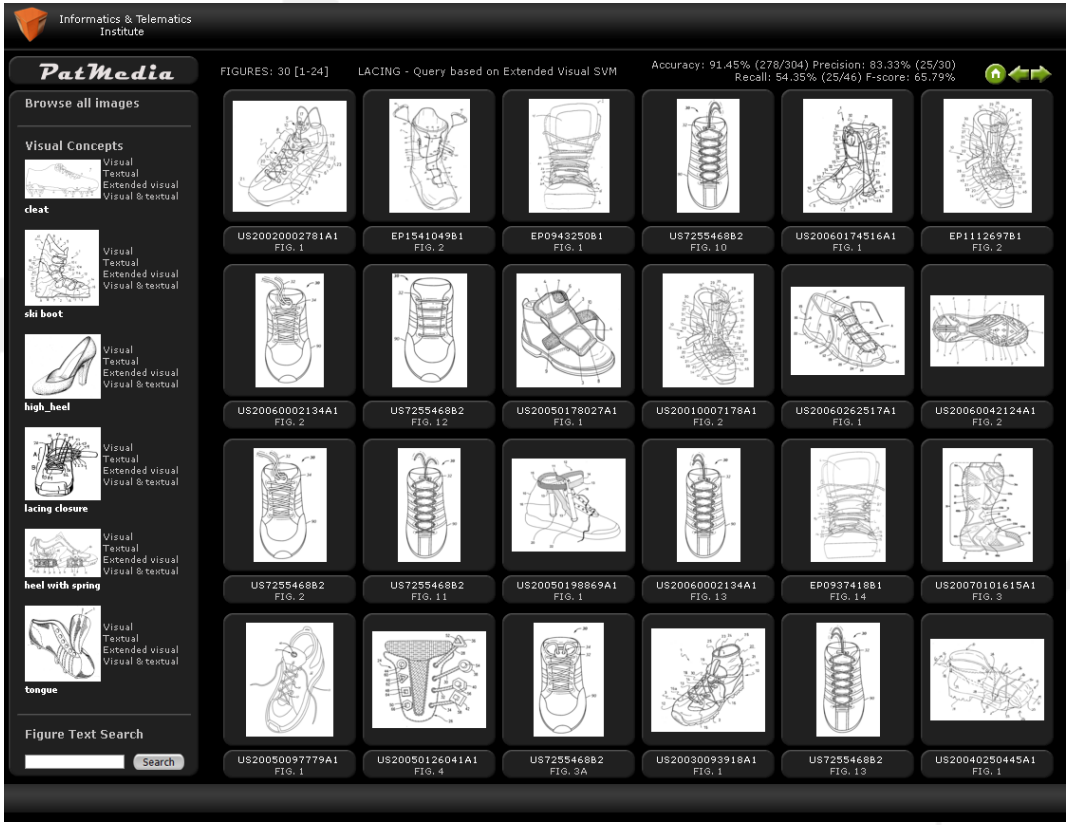
Visual
Textual
Textual + Visual

Visual extended

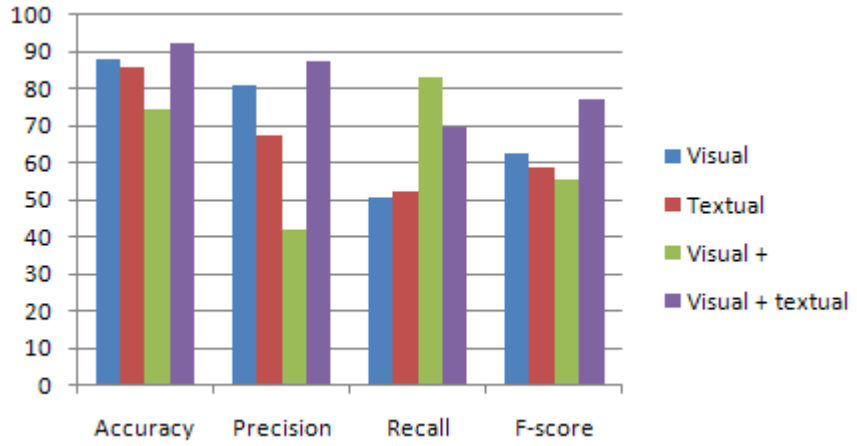
Max score output

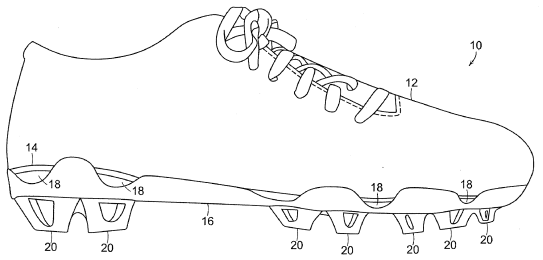
Time for the demo!

http://mklab-services.iti.gr/patmediac



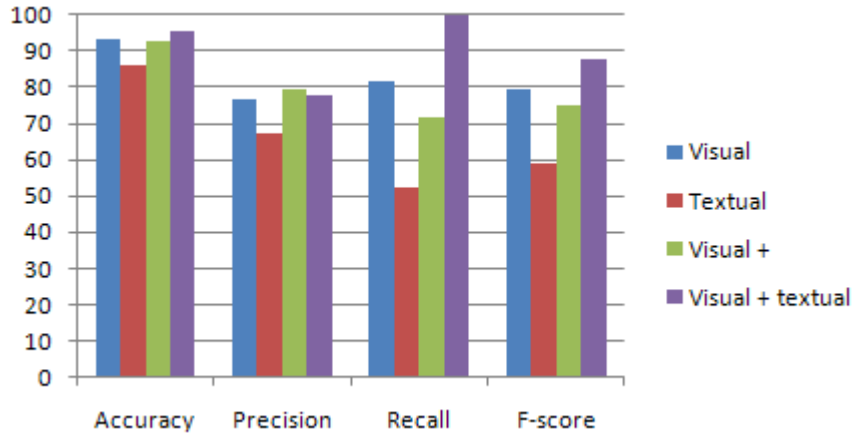
Results-Cleat

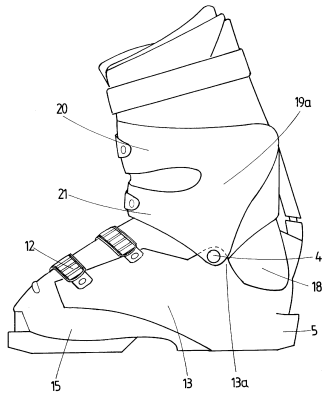




Features	Accuracy	Precision	Recall	F-Score
Visual	88,16%	81,08%	50,85%	62,5%
Textual	85,86%	67,39%	52,54%	59,05%
Visual ext.	74,34%	41,88%	83,05%	55,68%
Visual+Textual	92,11%	87,23%	69,49%	77,36%

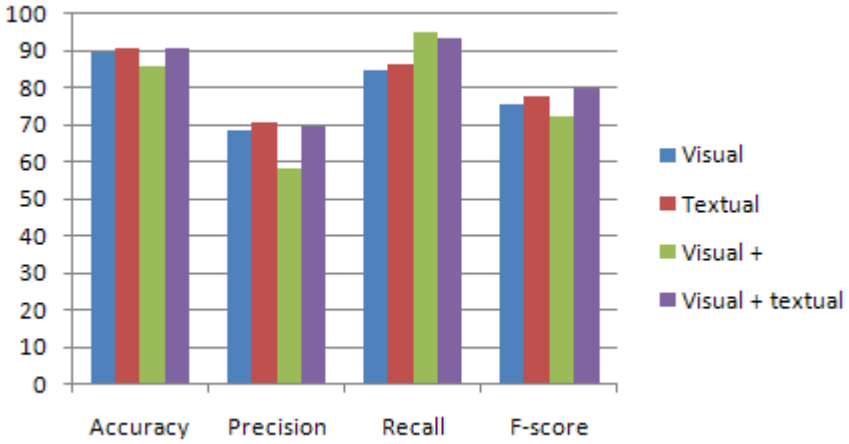
Results-Ski boot





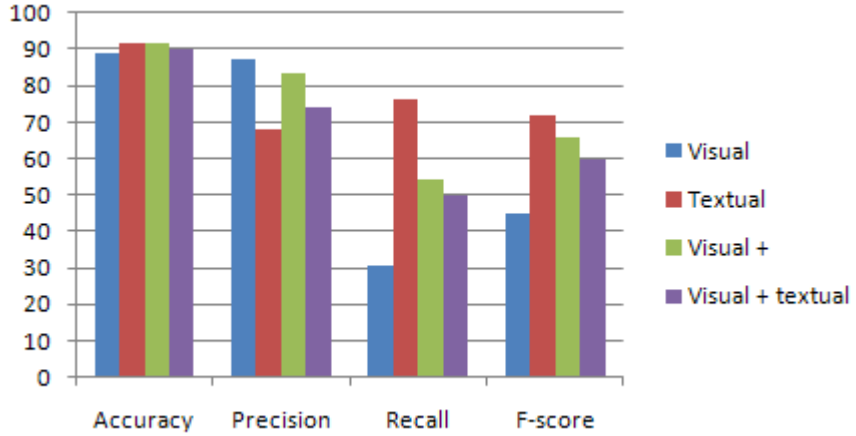
Features	Accuracy	Precision	Recall	F-Score
Visual	93,09%	76,92%	81,63%	79,21%
Textual	85,86%	67,39%	52,54%	59,05%
Visual ext.	92,43%	79,55%	71,43%	75,27%
Visual+Textual	95,39%	77,78%	100%	87,5%

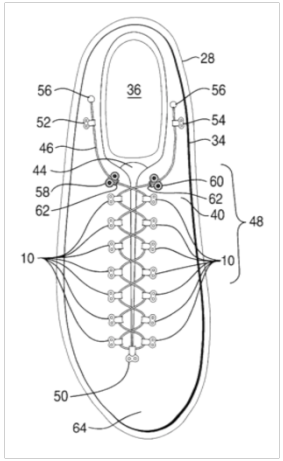
Results-High Heel



Features	Accuracy	Precision	Recall	F-Score
Visual	89,47%	68,49%	84,75%	75,76%
Textual	90,46%	70,83%	86,44%	77,86%
Visual ext.	85,86%	58,33%	94,92%	72,26%
Visual+Textual	90,79%	69,62%	93,22%	79,71%

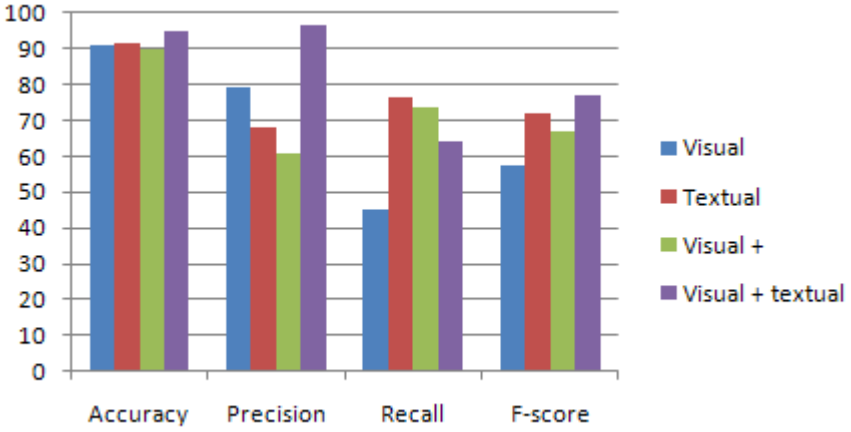
Results-Lacing Closure

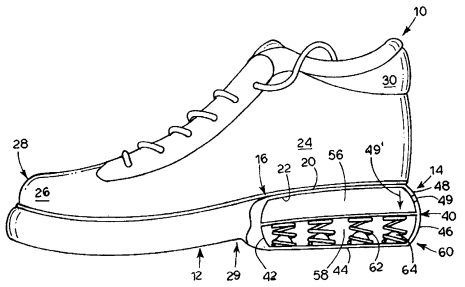




Features	Accuracy	Precision	Recall	F-Score
Visual	88,82%	87,5%	30,43%	45,16%
Textual	91,78%	68,09%	76,19%	71,91%
Visual ext.	91,45%	83,33%	54,35%	65,79%
Visual+Textual	89,8%	74,19%	50%	59,74%

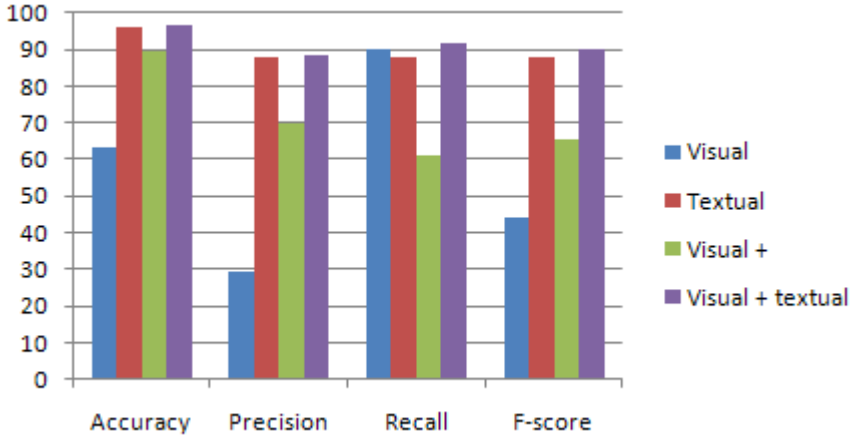
Results-Heel with spring

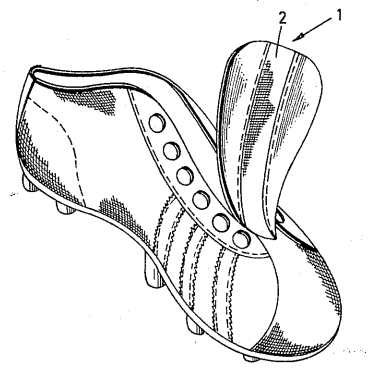




Features	Accuracy	Precision	Recall	F-Score
Visual	90,79%	79,17%	45,24%	57,58%
Textual	91,78%	68,09%	76,19%	71,91%
Visual ext.	89,8%	60,78%	73,81%	66,66%
Visual+Textual	94,74%	96,43%	64,29%	77,15%

Results-Tongue

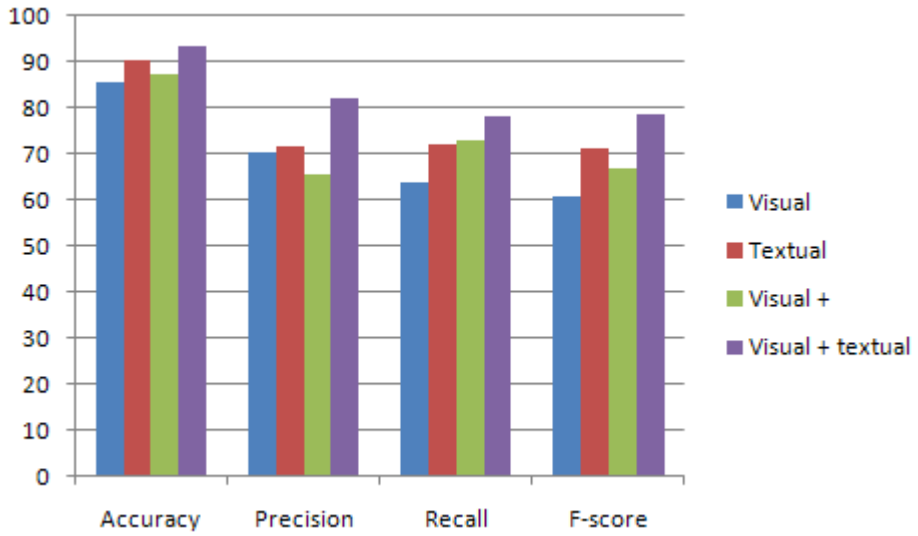




Features	Accuracy	Precision	Recall	F-Score
Visual	63,16%	29,14%	89,8%	44%
Textual	96,05%	87,76%	87,76%	87,76%
Visual ext.	89,47%	69,77%	61,22%	65,22%
Visual+Textual	96,71%	88,24%	91,84%	90%

Results

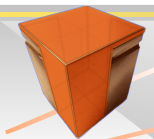
- Average Results for all concepts



Features	Accuracy	Precision	Recall	F-score
Visual	85,58%	70,38%	63,78%	60,7%
Textual	90,3%	71,59%	71,94%	71,25%
Visual ext.	87,22%	65,6%	73,13%	66,81%
Visual + Textual	93,25%	82,24%	78,14%	78,58%

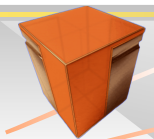
Conclusions

- 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.



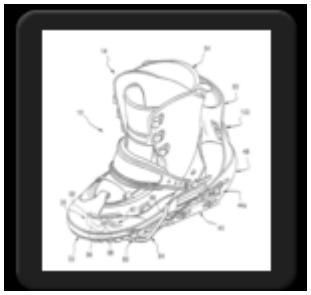
Future Work

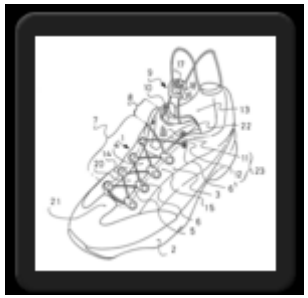
- Produce results for more concepts.
- Realize the same framework for bigger datasets and for different IPC classes.
- Test performance in the case of automatically segmented images (i.e. of lower quality).
- Combine more efficiently text and visual information
  - give more weight to visual or textual description depending on the concept type
- Investigate late fusion techniques.

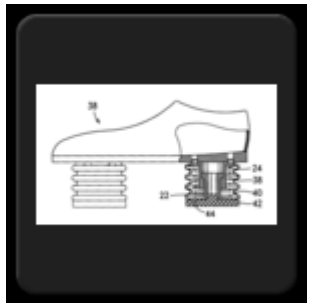


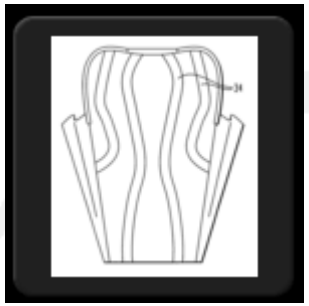
Feel free to test the demo!

http://mklab-services.iti.gr/patmediac









Thank you!

http://mklab.iti.gr