Concept-oriented labelling of patent images based on Random Forests and proximity-driven generation of synthetic data

Dimitris Liparas	Anastasia Moumtzidou	Stefanos Vrochidis	Ioannis Kompatsiaris
Information Technologies Institute Centre for Research and Technology Hellas Thermi-Thessaloniki, Greece dliparas@iti.gr	Information Technologies Institute Centre for Research and Technology Hellas Thermi-Thessaloniki, Greece moumtzid@iti.gr	Information Technologies Institute Centre for Research and Technology Hellas Thermi-Thessaloniki, Greece stefanos@iti.gr	Information Technologies Institute Centre for Research and Technology Hellas Thermi-Thessaloniki, Greece ikom@iti.gr

Abstract

Patent images are very important for patent examiners to understand the contents of an invention. Therefore there is a need for automatic labelling of patent images in order to support patent search tasks. Towards this goal, recent research works propose classification-based approaches for patent image annotation. However, one of the main drawbacks of these methods is that they rely upon large annotated patent image datasets, which require substantial manual effort to be obtained. In this context, the proposed work performs extraction of concepts from patent images building upon a supervised machine learning framework, which is trained with limited annotated data and automatically generated synthetic data. The classification is realised with Random Forests (RF) and a combination of visual and textual features. First, we make use of RF's implicit ability to detect outliers to rid our data of unnecessary noise. Then, we generate new synthetic data cases by means of Synthetic Minority Over-sampling Technique (SMOTE). We evaluate the different retrieval parts of the framework by using a dataset from the footwear domain. The results of the experiments indicate the benefits of using the proposed methodology.

1 Introduction

The vast number of patent documents submitted to patent offices worldwide calls for the need of advanced patent search technologies, which could deal effectively with the complexity and the unique characteristics of patents. The majority of existing patent retrieval techniques and search engines rely upon text, given the fact that the ideas and the innovations to be patented are described in text format in the claims and the disclosure parts of the patent. However, we should not overlook the fact that most of the patents include a drawings section, which contains figures, drawings and diagrams as a means to further describe and understand the patented inventions.

In the recent years, the Intellectual Property and Information Retrieval communities, motivated by the interest in patent image search, have directed their efforts towards the development of systems that have the ability to search in patents by considering both textual and visual information. Following the latest trends and challenges in image retrieval, the most recent studies in patent image search deal with concept extraction and classification using visual features (e.g. Csurka et al., 2011; Vrochidis et al., 2012). The concept extraction techniques involve the identification of images with common characteristics that fall into a specific semantic category or depict a specific concept. The motivation behind the interest in patent concept-based search is revealed by the following scenario presented in (De Marco, 2010): a patent searcher searches for a dancing shoe that incorporates a rotating heel with ball bearings; at first, the patent searcher recognises the main concepts of the invention (e.g. dancing shoe) and based on them keywords and relevant classification areas are defined. In many cases the important information is described with figures. Therefore, it would be important if the patent searcher could directly retrieve patents, which include figures depicting these concepts. The main obstacle of this approach is the need for a significant number of annotated images required during the training phase for developing models for each concept/category, something that is arduous and time-consuming (due to the specific nature of these images, it is not easy to retrieve training instances from the web).

To deal with the aforementioned restriction, we present an approach for concept extraction from patent images with the ability to supplement a small manually annotated dataset by means of automatic synthetic data cases generation. The proposed methodology is based on a supervised machine learning framework using Random Forests (RF) trained with textual and visual features. RF's advantage of handling multiclass classification tasks directly eliminates the need to develop a classification model for each concept separately. Moreover, its outlier detection technique carries out a suitable pre-processing of the data. The main contribution and the research objective of this paper is the examination of concept extraction based on multimodal classification, coupled with RF construction driven by synthetic data and outlier elimination. While the research works up to date apply Synthetic Minority Over-sampling Technique (SMOTE) for the purpose of overcoming imbalanced-related problems, the proposed approach extends the application of SMOTE to the generation of synthetic cases in an already balanced dataset. To the best of our knowledge, there isn't any relevant literature concerning the application of SMOTE to multiclass datasets that are balanced but contain a relatively small amount of training instances per class (which is the case in this study).

The rest of the paper is organised as follows: In Section 2 we provide the theoretical background of our study. In Section 3, the related work is presented. The feature extraction process and the architecture of the proposed framework are analysed in Sections 4 and 5, respectively. Section 6 describes the conducted experiments, as well as the results. Finally, concluding remarks are provided in Section 7.

2 Theoretical background

Random Forests (RF) is an ensemble learning method for classification and regression (Breiman, 2001). Its inherent ability to learn multiclass classification problems (without the need to convert the multiclass problem into a set of binary classification problems) makes it one of the most attractive machine learning algorithms. The fundamental idea of the methodology is the construction of a multitude of decision trees. RF operates on two sources of randomness. Firstly, each decision tree is grown on a different bootstrap sample drawn randomly from the training data. Secondly, at each node split during the construction of a decision tree, a random subset of *p* variables is selected from the original variable set and the best split based on these *p* variables is used. For predicting an unknown case, the predictions of the trees constituting the RF are aggregated (majority voting for classification / averaging for regression). For a RF consisting of *N* trees, the equation for predicting the class label *l* of a case *y* through majority voting is the following:

$$l(y) = argmax_c(\sum_{n=1}^{N} I_{h_n(y)=c}) \tag{1}$$

where *I* the indicator function and h_n the *nth* tree of the RF.

Among other things, RF can provide an internal estimate of its generalisation error. This is achieved by the out-of-bag (OOB) error estimate. For each tree that is constructed, only 2/3 of the original data cases are used in that particular bootstrap sample. The rest 1/3 of the instances (OOB data) are classified by the constructed tree and therefore, used for testing its performance. The OOB error estimate is the averaged prediction error for each training case *y*, using only the trees that do not include *y* in their bootstrap sample. Moreover, RF has a built-in mechanism for detecting outliers. Within the RF context, cases whose proximities to all other cases in the data are generally small can be considered outliers (Breiman, 2001). When a RF is constructed, all the training cases are put down each tree and their proximity matrix is computed, based on whether pairs of cases end up in the same terminal node of a tree. From this proximity matrix, an outlier measure for each case is derived. Cases whose outlier measure values exceed a specified threshold are detected as outliers. For a more thorough analysis of the core concepts of RF, see (Breiman, 2001).

Synthetic Minority Over-sampling Technique (SMOTE) is an approach for constructing efficient classifiers from imbalanced datasets (Chawla et al., 2002). A dataset can be defined as imbalanced if its classes are not evenly represented. The basic notion of the technique is the synthetic generation of new minority class examples, based on the nearest neighbours of these cases, coupled with the under-sampling of the majority class cases (Chawla et al., 2002).

3 Related work

Since our proposed framework deals with patent image classification, we report previous work related to patent image concept extraction and to the classification methodologies involved in this study.

3.1 Patent image search and classification

The first attempts in patent image search were based on the extraction of visual low level features with a view of retrieving visually similar images based on the query by visual example paradigm. Within this context, several systems have been developed, including PATSEEK (Tiwari and Bansal, 2004) and PatMedia image search engine (Vrochidis et al., 2010).

More recently, research in patent image search moved towards concept extraction and classification. The two main approaches followed for concept generation in multimedia content are: “content-based” and “text-based”. The content-based analysis uses visual low-level features to represent the multimedia content. In such a work (Csurka et al., 2011) the authors extract SIFT-like local orientation histograms and they build visual vocabularies specific to patent images using Gaussian mixture model (GMM). Then the images are represented by Fisher features and linear classifiers are employed for the categorisation. On the other hand, text-based representation uses the indexing of media according to text that can be associated to it, such as titles or descriptions in associated metadata files. Although text-based representation can be considered as reliable, it depends on the existence and the quality of the annotations. Finally, other recent works consider both visual and textual information. For example, in (Vrochidis et al., 2012) the authors propose a supervised machine learning framework to extract semantic concepts from patent images by combining visual and textual information. Although the aforementioned studies deal with patent image concept extraction and classification, none of them considers the use of synthetically generated data to leverage the classification performance.

3.2 Random Forests/SMOTE

Over the years, RF has been successfully applied to a wide range of disciplines. More specifically, several studies dealing with image classification can be found in the relevant literature (see for example (Bosch et al., 2007)). In this domain, a number of modifications of the RF algorithm have been proposed (Moosmann et al., 2008; Xu et al., 2012). In addition, there has been some research addressing the outlier detection mechanism provided by RF (Zhang and Zulkernine, 2006).

Regarding SMOTE, many studies dealing with class imbalance problems have used this method to overcome such issues. Among others, (Wang, 2008; Gao et al., 2011) can be listed. Moreover, several improvements of the original algorithm have been introduced (Chawla et al., 2003; Wang et al., 2006).

In general, RF, as many other popular machine learning methods, needs a large amount of training data, in order to achieve good performance and be able to generalise to new, unknown instances. For any given classification problem, obtaining large datasets that contain existing training examples is not always feasible, therefore a need to create new synthetic data arises. Although existing applications of SMOTE are dealing with balancing imbalanced data (Wang, 2008; Gao et al., 2011), in this work we use SMOTE with the sole purpose of generating synthetic cases, in order to enrich and improve the training procedure of RF in a patent image classification framework.

4 Feature extraction from patent images

In this Section we briefly describe the extraction of visual and textual features from the patent images.

4.1 Visual features

The extraction of global concepts requires the employment of global visual features, which can capture the special characteristics of patent images (i.e. they are mostly black and white and depict technical drawings). Given the fact that general case image representation features consider colour and texture, which are absent in most of patent images, it is evident that we need to apply an algorithm that takes into account the geometry and the pixel distribution of these images. To this end, we employ the

Adaptive Hierarchical Density Histograms (AHDH) as visual feature vectors, due to the fact that they have shown discriminative power between binary complex drawings (Sidiropoulos et al., 2011).

The AHDH feature vector is generated based on the following steps. First, the algorithm involves a pre-processing phase for noise reduction, coordinate calculation and normalisation. Then, the first geometric centroid of the image plane is calculated and the image area is split into four regions based on the position of this centroid. In the following, the feature vector is initialised by estimating the distribution of the black points in each region. This procedure is repeated in a recursive way for a manually specified number of iterations (e.g. Fig. 1), and after each iteration, the feature vector is updated. This non-segmentation point-density orientated technique combines high accuracy at low computational cost as it represents the image with a low dimension feature vector (i.e. around 100 features).

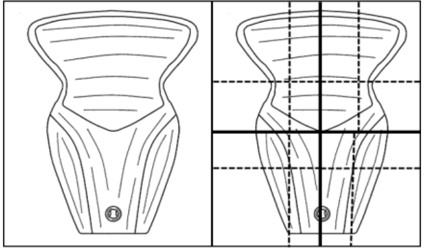

Fig. 1. The image is iteratively split to new regions based on the geometric centroids.

4.2 Textual features

Each figure of the patent document is linked to a description and caption found within the text. In order to exploit these textual descriptions, we apply a bag of words approach to model each figure with a vector. The bag of words model is a simplifying assumption used in natural language processing and information retrieval. In this model, a text (such as a sentence or a document) is represented as an unordered collection of words, disregarding grammar and even word order.

To generate such a vector we define a lexicon, which includes the most frequent words of this dataset. Then for each figure and based on the associated description we calculate a weight for each word included in the lexicon. The textual annotations are stemmed using the Porter stemming algorithm and the frequent stop words (e.g. and, so, etc.) are removed. The weight of each term is calculated with the well-established metric tf-idf (term frequency multiplied with the inverse document frequency).

5 Proposed methodology framework

The flowchart of the proposed concept extraction and classification framework (training phase) is depicted in Fig. 2. Next, the different steps and components of the framework are described in detail.

First, the **patent images and the captions associated to them are extracted** from the patent and in the following step the **visual and textual features are generated**, according to the procedures described in Section 4. In this approach we treat each modality's features independently. Thus, two different feature vectors (one for each modality) are formulated.

In the training phase, the feature vectors from each modality serve as input for the **construction of a RF** (section 2), from which we proceed to the **detection of possible outliers**. The latter is achieved in the following way: for each RF, the corresponding dataset's training cases are put down each tree. If a pair of cases end up in the same terminal node of a tree, their proximity is increased by 1. This is repeated for every pair of cases and all trees in the RF. In order to obtain the final proximity values we normalise them (divide them by the number of trees). Thus, if a dataset consists of N cases, a $N \times N$ proximity matrix is derived. From this proximity matrix a measure that indicates the outlierness of each case is computed. In general, since the RF algorithm is based on randomisation and in order to obtain robust and reliable estimations about potential outliers, we suggest that the RF construction for the outlier detection and elimination step is repeated several times for each modality and the resulting outlier measure values from the constructed RFs are averaged. In this way, the randomisation factor is minimised and the outliers can be detected with more confidence. We note that we opted to follow this approach. The cases that are identified as outliers are eliminated from further processing.

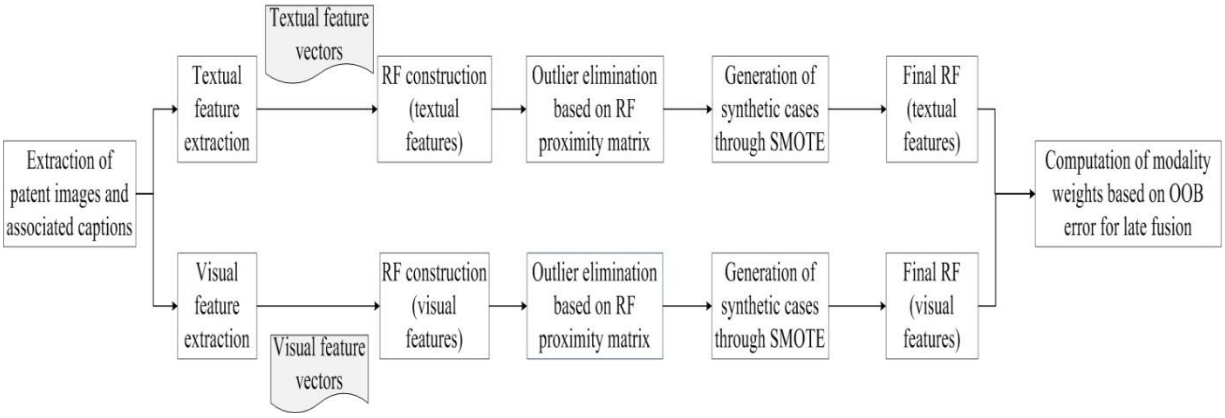
Textual feature vectors
Textual feature extraction
RF construction (textual features)
Outlier elimination based on RF proximity matrix
Generation of synthetic cases through SMOTE
Final RF (textual features)
Extraction of patent images and associated captions
Computation of modality weights based on OOB error for late fusion
Visual feature extraction
RF construction (visual features)
Outlier elimination based on RF proximity matrix
Generation of synthetic cases through SMOTE
Final RF (visual features)
Visual feature vectors

Fig. 2. Flowchart of proposed patent image classification framework

Next, the over-sampling procedure of SMOTE is employed, in order to **artificially generate new cases** and supplement the existing ones. The resulting larger datasets can lead to a better and more efficient RF training. It is important to note that the classes of the dataset used in this study are balanced in the first place. Therefore, we apply SMOTE not for balancing the classes of the training data, which was its main application to date (Section 2), but for introducing new training cases. According to (Chawla et al., 2002), SMOTE over-samples each case by introducing synthetic examples along the line segments joining a number (the number depends on the amount of over-sampling required) of that case's nearest neighbours. The over-sampling procedure is applied to each modality's dataset and to each concept separately. In this way the final datasets are created and correspondingly, the **final RFs for the textual and visual features are constructed**.

Finally, for the **formulation of the final RF predictions, a late fusion strategy is applied** as follows: from the OOB error estimate (for the entire data set) of each modality's RF, the corresponding OOB accuracy values are computed. These values are normalised (by dividing them by their sum) and serve as weights for the two modalities. During the testing phase, when the RF predicts a case, it outputs probability estimates per class for that case. The probability outputs P_t and P_v from the textual and visual RFs respectively are multiplied by their corresponding modality weights W_t and W_v and summed, in order to produce the final RF predictions as in the following equation:

$$P_{fused} = W_t P_t + W_v P_v \tag{2}$$

6 Experimental design - Results

6.1 Dataset description – Experimental setup

The dataset[1] was manually extracted from around 300 patents. It contains around 1000 patent images depicting parts of footwear. The feature vectors that were generated (Section 4) for this dataset consist of 100 visual and 250 textual features. With the help of professional patent searchers we selected the following 8 concepts for this domain: cleat, ski boot, high heel, lacing closure, heel with spring, tongue, toe caps and roller skates (Vrochidis et al., 2012). The procedure of associating the patent images with the figure text descriptions was carried out manually. This was done in order to acquire quality data and consequently, draw safer conclusions on the concept extraction method. The images were manually annotated with the support and advice of professional patent searchers.

For our experiments, the dataset was randomly split into training and test sets. We kept 2/3 of the images for training purposes, whereas the rest (1/3) were used as test set, in order to estimate the classification scheme's performance. We note here that because of the fact that RF provides an internal estimate of its performance on cases that do not participate in its training procedure (the OOB error estimate), no cross-validation was required. Moreover, since SMOTE is applied during the training phase, it is important to mention that the test set contains only real (not synthetic) data.

[1] Available for download at http://mklab.iti.gr/files/concepts-patent_images.rar

Regarding the parameters of the methods involved in the experiments, we selected and applied the following setting: The number of trees used for the construction of each RF was set based on the OOB error estimate. After conducting several experiments and gradually increasing the number of trees, we noticed that the OOB error estimate was stabilised after using 1000 trees and no longer improved. Hence, the number of trees was set to 1000. Moreover, for each node split during the growing of each tree, the number of the subset of variables used to determine the best split was set to $\sqrt{k}$, where k is the total number of features of the dataset (according to (Breiman, 2001)). Concerning the RF outlier detection and elimination procedure, (Breiman, 2003) states that a case can be considered an outlier if its outlier measure value is higher than 10. We note that after choosing this configuration, approximately 2% of the textual modality's cases were detected as outliers and discarded, keeping the rest of the cases for further processing, while for the visual modality no outliers were detected. Finally, the SMOTE oversampling rate for each concept in both modalities datasets was set to 500%, i.e. for each case 5 new synthetic cases were generated, based on this case's nearest neighbours.

6.2 Results

In order to evaluate the performance of the proposed methodology, we computed the precision, recall and F-score measures for each concept, along with their corresponding macro-averaged values. Table 1 summarises the test set results from the application of RF to the initial dataset, without any outlier deletion and without the use of SMOTE. The results refer to each modality separately, as well as to their fused (according to the OOB error estimates) output scores (Visual + Textual). Since the F-score takes into account precision and recall simultaneously, we consider it the most important metric for the evaluation of the results. Moreover, since we are handling this classification problem in a multiclass manner, we are more interested in the macro-average value of F-score. The results verify the notion mentioned in Section 1 that textual data is a very reliable source of information for the patent retrieval, since the textual modality achieves a macro-averaged F-score value of 73.7%, compared to 67.4% for the visual modality. However, the late fusion of the outputs of each modality's RF provides us with improved results compared to the ones provided by each single modality, as evident from the corresponding macro-averaged F-score value (81.8%). This suggests that the visual and textual modalities can complement each other in the patent image classification task. If we observe the results for each concept independently, we notice that the textual features do not outperform the visual ones only for the "Ski boot", "High heel" and "Lacing" concepts. On the other hand, the fused results are better than the corresponding visual and textual results for every concept, with only the exception of "Tongue", where the textual features achieve the same performance (89.9% for the F-score).

Concepts	**Visual**			**Textual**			**Visual + Textual**		
	Prec.	Rec.	F-score	Prec.	Rec.	F-score	Prec.	Rec.	F-score
Cleat	73.3%	55.9%	63.4%	70.1%	67.8%	68.9%	82.4%	79.6%	80.9%
Ski boot	94.4%	69.4%	80%	74.1%	81.6%	77.6%	76.2%	91.8%	83.2%
High heel	63.6%	83%	71.9%	51.7%	76.2%	61.5%	76.4%	93.2%	83.9%
Lacing	51.5%	71.7%	59.9%	62.8%	47.8%	54.2%	76.3%	63%	69%
Spring	70.5%	73.8%	72.1%	89.6%	61.9%	73.2%	90.6%	69%	78.3%
Tongue	76.6%	46.9%	58.1%	88.2%	91.8%	89.9%	88.2%	91.8%	89.9%
Toe caps	70.6%	55.8%	62.2%	83.8%	72.1%	77.5%	82.9%	79.1%	80.9%
Roller	64.3%	80.6%	71.5%	89.1%	85.1%	87%	90.6%	86.5%	88.5%
Macro-average	**70.6%**	**67.1%**	**67.4%**	**76.2%**	**73%**	**73.7%**	**82.9%**	**81.7%**	**81.8%**

Table 1. Precision, recall and F-score test set results (without outlier deletion and without SMOTE)

In Table 2 we report the test set results from the RF application to the dataset after the deletion of the detected outliers and the use of SMOTE. While a minor degradation for the F-scores of some of the concepts for both modalities (compared to the RF application to the original dataset) is obvious, it seems that in average RF has benefited from the outlier elimination procedure and the generation of new cases through SMOTE, as the macro-averaged F-scores have been improved (69.5% for the visual and 75.7% for the textual features). In this case, the visual-based classification performs better than the textual-based one only for the "High heel" and "Lacing" concepts. The fusion of the modalities out-performs each one of them (84.2%) and moreover, there is a 2.4% improvement compared to the fused results for the initial dataset. Finally, in Figs. 3 and 4 the first 6 results for the “Ski boot” and “Tongue” concepts (respectively), using the final dataset, are presented. The precision achieved for this set of results is 83.3% (5/6) for the “Ski boot” and 100% (6/6) for the “Tongue” concept.

Concepts	**Visual**			**Textual**			**Visual + Textual**		
	Prec.	Rec.	F-score	Prec.	Rec.	F-score	Prec.	Rec.	F-score
Cleat	66.1%	66.1%	66.1%	79.2%	71.2%	75%	89.1%	83.1%	85.9%
Ski boot	85.7%	73.5%	79.1%	77.7%	85.7%	81.5%	80.4%	83.7%	81.9%
High heel	68.6%	81.4%	74.4%	76.9%	67.8%	72%	80.6%	84.7%	82.6%
Lacing	50%	76.1%	60.3%	42.4%	60.9%	50%	67.3%	76.1%	71.4%
Spring	68.1%	71.4%	69.7%	73.9%	81%	77.3%	90.2%	88.1%	89.1%
Tongue	78.3%	59.2%	67.4%	86.3%	89.8%	88%	88.2%	91.8%	89.9%
Toe caps	72.2%	60.5%	65.8%	90.6%	67.4%	77.2%	89.7%	81.4%	85.3%
Roller	74.2%	73.1%	73.6%	90%	80.6%	85%	90.5%	85.1%	87.7%
Macro-average	**70.4%**	**70.1%**	**69.5%**	**77.1%**	**75.5%**	**75.7%**	**84.5%**	**84.2%**	**84.2%**

Table 2. Precision, recall and F-score test set results (with outlier deletion and SMOTE)

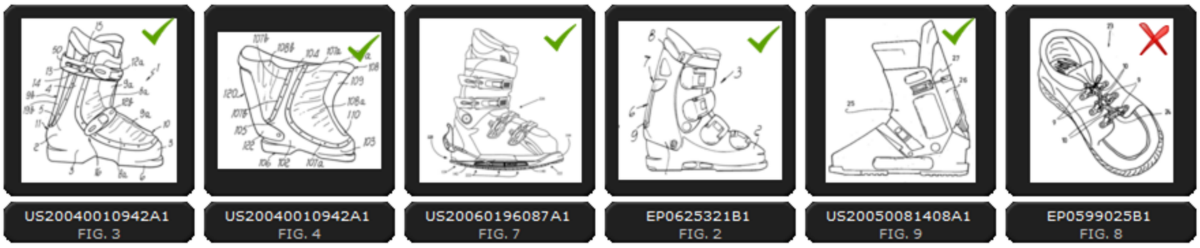

Fig. 3. Results for the “Ski boot” concept. The green tics indicate the correct results.

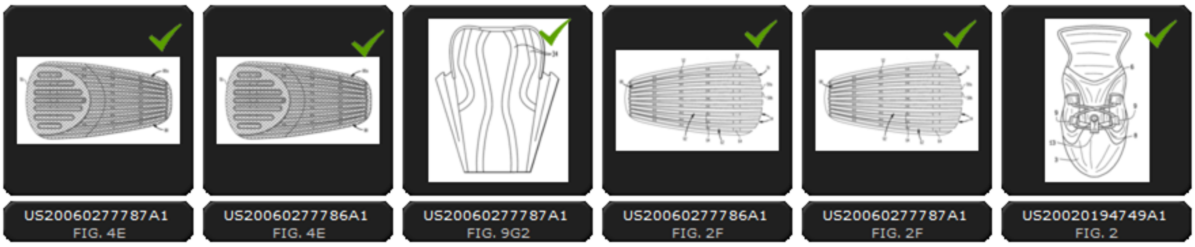

Fig. 4. Results for the “Tongue” concept. The green tics indicate the correct results.

7 Conclusions

In this study a concept extraction and multimodal classification framework for patent images trained with synthetic data, is introduced. The implicit ability of RF to deal effectively with multiclass classification problems and to perform a suitable filtering of the data, according to how well the dataset’s instances fit in its construction procedure, coupled with the benefits of using an over-sampling technique such as SMOTE, provide a final result that leads to enhanced classification performance.

The application of this framework could augment existing (mainly text-based) patent search systems. Although the framework has been tested with a limited set of concepts, the methodology based on RF is scalable and the application of SMOTE minimises the need for training data. In addition the application of the concept extraction methodology to patents that belong to the same IPC (International Patent Classification) class and/or groups will allow for targeting only a specific set of concepts (relevant to the corresponding IPC class). The concept-based retrieval functionality will enable patent examiners to search in patent figures based on their visual content and therefore speed up and improve the performance of patent search tasks for patent invalidation and competitive intelligence research.

Our recommendations for future work include the testing of different parameter settings than the one used in this study and the evaluation of their performance, the expansion of the experiments with a very large set of concepts and finally, the investigation of alternative multimodal fusion approaches, such as the one presented in (Roller and Schulte im Walde, 2013).

Acknowledgment: This work was supported by MULTISENSOR project (FP7-610411).

Reference

Anna Bosch, Andrew Zisserman, and Xavier Munoz. 2007. *Image classification using random forests and ferns*. In ICCV, 1-8.

Leo Breiman. 2001. *Random Forests*. In Machine Learning, 45(1): 5-32.

Leo Breiman. 2003. *Manual – Setting up, using and understanding random forests v4.0*. http://oz.berkeley.edu/users/breiman/ Using_random_forests_v4.0.pdf.

Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. 2002. *SMOTE: Synthetic Minority Over-Sampling Technique*. Journal of Artificial Intelligence Research, 16: 321-357.

Nitesh V. Chawla, Aleksandar Lazarevic, Lawrence O. Hall, and Kevin W. Bowyer. 2003. *SMOTEBoost: Improving prediction of the minority class in boosting*. In 7th European Conference on Principles and Practice of Knowledge Discovery in Databases, 107–119.

Gabriela Csurka, Jean-Michel Renders, and Guillaume Jacquet. 2011. G. *XRCE's Participation at Patent Image Classification and Image-based Patent Retrieval Tasks of the Clef-IP 2011*. In: Proceedings of CLEF 2011, Amsterdam

Dominic De Marco. 2010. *Mechanical Patent Searching: A Moving Target*. Patent Information Users Group (PIUG), Baltimore, USA

Ming Gao, Xia Hong, Sheng Chen, and Chris J. Harris. 2011. *A combined SMOTE and PSO based RBF classifier for two-class imbalanced problems*. Neurocomputing, 74:3456–3466.

Frank Moosmann, Eric Nowak, and Frederic Jurie. 2008. *Randomized clustering forests for image classification*. IEEE Transactions on PAMI, 30(9): 1632-1646.

Stephen Roller and Stephen Schulte im Walde. 2013. *A multimodal LDA model integrating textual, cognitive and visual modalities*. In Proceedings of the 2013 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, 1146–1157, Seattle, Washington, USA

Panagiotis Sidiropoulos, Stefanos Vrochidis, and Ioannis Kompatsiaris. 2011. *Content-based binary image retrieval using the adaptive hierarchical density histogram*. *Pattern Recognition Journal*, 44(4):739–750.

Avinash Tiwari and Veena Bansal. 2004. PATSEEK: *Content Based Image Retrieval System for Patent Database*. In: Proceedings International Conference on Electronic Business, Beijing, China

Stefanos Vrochidis, Symeon Papadopoulos, Anastasia Moumtzidou, Panagiotis Sidiropoulos, Emanuele Pianta, and Ioannis Kompatsiaris. 2010. *Towards Content-based Patent Image Retrieval; A Framework Perspective*. *World Patent Information Journal*, 32(2):94-106.

Stefanos Vrochidis, Anastasia Moumtzidou, and Ioannis Kompatsiaris. 2012. *Concept-based Patent Image Retrieval. World Patent Information Journal*, 34(4):292-303.

Juanjuan Wang, Mantao Xu, Hui Wang, and Jiwu Zhang. 2006. *Classification of imbalanced data by using the SMOTE algorithm and locally linear embedding*. In 8th International Conference on Signal Processing, 3:16–20.

He-Yong Wang. 2008. *Combination approach of SMOTE and biased-SVM for imbalanced datasets*, Proc. of the IEEE Int. Joint Conf. on Neural Networks, IJCNN 2008, Hong Kong (PRC), 22-31.

Baoxun Xu, Yunming Ye, and Lei Nie. 2012. *An improved random forest classifier for image classification*. In Information and Automation (ICIA), 2012 International Conference on IEEE, 795-800.

Jiong Zhang and Mohammad Zulkernine. 2006. *A Hybrid Network Intrusion Detection Technique Using Random Forests*. In Proceedings of IEEE First International Conference on Availability, Reliability and Security (ARES'06).