How to Compare Search Engines in CBIR?

Tatiana Jaworska
Systems Research Institute,
Polish Academy of Sciences
6 Newelska Street, 01-447 Warsaw, Poland
Tatiana.Jaworska@ibspan.waw.pl

***Abstract*—At present a great deal of research is being done in different aspects of Content-Based Image Retrieval (CBIR) of which the search engine is one of the most important elements. In this paper we cover the state-of-the-art techniques in CBIR according to the aims of retrieval and matching techniques. The issue we address is the analysis of search engines reducing the 'semantic gap'. The matching methods are compared in terms of their usefulness for different user's aims. Finally, we compare our search engine with Google's and the SIFT method.**

Keywords— CBIR; search engine; SIFT; image query; image retrieval.

I. INTRODUCTION

In the last decade, the availability of large image datasets and search engines has increased tremendously. It is obvious that there is no universal CBIR system for finding all images and the spectrum of available systems ranges from the general purpose ones, like Google, to very narrowly specialized ones, like those found in medicine or astronomy. This multitude has necessitated a review in order to find the most suitable system for the user's purpose. The basic list of search engines can be obtained on the Internet [1].

Early on search engines used low-level features, such as colour, shape, texture information and annotations to retrieve similar images. This approach is still popular, but although many algorithms have been developed, they cannot adequately model image semantics and have many limitations when dealing with the vast resources of image databases. A survey on low-level image feature extraction in CBIR systems can be found in [2].

Hence, currently, the predominant engine categories are based on [3]:

- object ontology introduced to define high-level concepts,
- bag-of-visual-words (BoW), stemming from the text analysis,
- object retrieval using SIFT and its modification methods,
- relevance feedback (RF) implemented into a retrieval loop for continuous learning about users' intention,
- a semantic template (ST) defined to support high-level image retrieval,
- the information covering the visual content of images and the textual description received from the Web for online image retrieval,
- combining visual properties of selected objects (or a set of relevant visual features), spatial or temporal relationships of graphical objects [4], [5], with semantic properties [6], [3].

The main contribution of this paper is the comparison of high-level semantic CBIRs with our new search engine which takes into account the kind and number of objects, their features, together with different spatial location of segmented objects in the image. Our search engine uses the GUI which enables the user to construct their own query image from the segmented objects.

The rest of the paper is organized as follows: Section II provides the aims of our search engine construction, section III surveys the matching techniques and describes our search engine in comparison with the others, with some implementation details. Section IV presents some results obtained from our engine and compares them with Google's and the SIFT method.

II. AIMS OF THE SEARCH ENGINE CONSTRUCTION

CBIR systems should meet the user's diverse requirements depending on the interest domain and the particular need. The user has to answer some questions of which the first and foremost is how to define their goal; do they want to construct a new CBIR system from scratch or build it on their existing image collections, for example, art collections, medical images, scientific databases or generally, the World Wide Web.

The next question which is inextricably connected with later selection criteria is whether there is a necessity of retrieval of whole images, object groups or possibly video fragments.

Another piece of required information is whether the annotations are assigned to the images in a DB. The answer to these problems will determine a single matching mechanism, listed above, as more efficient than the others.

Some other users need to put some order in their messy collection, while others want to find one object in many pictures, e.g. a face in an airport video, etc.

In the next subsection we will present advantages and disadvantages of the above-mentioned search engine categories.

978-1-4673-8460-5/16/$31.00 ©2016 IEEE

III. MATCHING TECHNIQUES

A. Object Ontology

Generally speaking, ontologies define the concepts and relationships used to describe and represent an area of knowledge. Ontology makes it possible to model the semantics contained in images, such as objects or events. It provides, in a formal way, mutual understanding in a specific domain between humans and computers. Hence, ontology represents knowledge in a hierarchical structure which is used to describe and organize an image collection and it also shows the relation between these images.

In the early approaches high-level concepts were described using the intermediate-level descriptors of the object's ontology. These descriptors were automatically mapped from the low-level features calculated for each region in the database, thus allowing the association of high-level concepts and potentially relevant image regions [7]. Later, ontology was employed to spatial relationships in images, such as connectivity, disjoint, meet, adjacency, overlap, cover, or inside. But the image was divided into 3x3, 5x5 or 9x9 windows instead of separate objects [8].

For ontological DBs, the Web Ontology Languages (OWL), as a family of knowledge representation languages, have been constructed for authoring ontologies characterized by formal semantics. In ontological approaches, the semantic information contained in image annotation is taken into account in order to reduce the number of feature vectors and to decrease the processing time.

Doulaverakis [9] presented a hybrid system devoted to the retrieval of real cultural heritage collections. A proposed search engine was capable of retrieving images based on their keyword annotation with the help of an ontology, or based on the image content to find similar images, or on both these strategies. This engine was composed of two different subsystems, a low-level image feature analysis with a retrieval system and a high-level ontology-based metadata structure. Both subsystems could co-operate during the evaluation of a single query in a hybrid way.

At present, applications use some separate ontologies. For example, Allani et al. [10] defined an image content ontology O_c with a set of image concepts, a meta-data ontology O_m addressing the surrounding textual information about an image and a visual feature ontology O_F (see Fig. 1) with a set of low-level image features. When a query image is introduced, image annotation is processed in order to extract concepts and use them to select relevant features to apply during the retrieval process. Query images are classified given their content into 6 classes. On each class of query images 7 retrieval strategies are performed given feature categories.

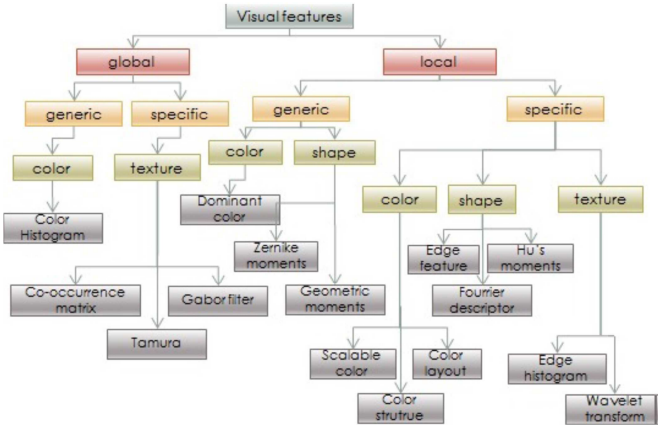
Ontology is also a method for organizing extra large-scale image collections, like the ImageNet dataset, created at Stanford University [11].

There are some advantages of an ontology:

- its application bridges the semantic gap;
- there is a special language for the user to ask a question;
- ontology-based algorithms are easy to implement and are suitable for applications with simple semantic features.

The disadvantage is the necessity of preparing a special DB and annotating the introduction.

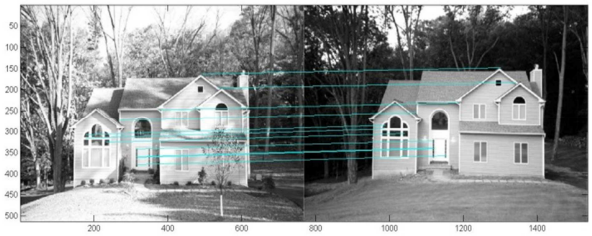
Fig. 1. Visual feature ontology **[10]**.



B. Object Retrieval Using SIFT

The scale invariant feature transform (SIFT) was introduced by Lowe *[12]*, *[13]* to identify objects in two images, even if these objects were cluttered or partially covered. Additionally, the SIFT feature descriptor helps matching objects which differ in scale, colour, or orientation.

Fig. 2. Point-to-point correspondence found by the SIFT descriptors



An object in a query image is identified in a second image by extracting features from both images. Possible matching feature vectors are found using the Euclidean metric. From all the potential matches, only a subset of key points is selected. Each key point is characterised by four parameters: x, y being the centre coordinates of the circular region whose r is its scalable radius and angle θ determines one of eight main directions. Based on these features the good matches are filtered out. In order to quickly determine clusters of key points a hash table is implemented, employing the generalized Hough transform. The clusters whose features agree on an object and its location undergo additional model verification in detail, whereas the weak matched clusters are rejected. Eventually, the Bayesian probability analysis points to the number of probable true and false matches which shows the existence of an image object. The object matches that pass all these tests can be identified as correct with high confidence.

978-1-4673-8460-5/16/$31.00 ©2016 IEEE

The basic SIFT advantage is its invariance to uniform scaling, orientation, as well as of affine distortion and changes of illumination. This property suggested that this method retrieves all images containing a specific object, even in a large scale image dataset, when that object is given as a query by example (QBE).

Hence, SIFT needs the query-by-example, but in some situations it may be difficult to provide, for instance, when we have an image in our mind but it is difficult to find it as a QBE and additionally, we do not need a whole collection of similar images.

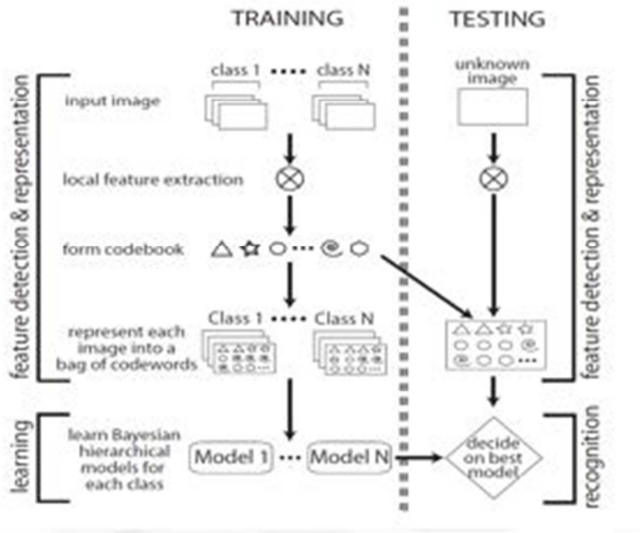
SIFT's additional advantage is the fact that it solves the problem of searching for disparity, independently of the issue of epipolar lines in stereovision. The example of point-to-point correspondence is presented in Fig. 2.

The local feature descriptors have undergone many modifications recently, for example, RootSIFT [14], RIFT [15], or BRIFT [16], etc.

C. Bag of Visual Words

A simple method of image classification is to treat them as a set of segments, describing only their appearance and ignoring their spatial layout which is very important in image representation. Similar approach have been successfully employed in the text collections to analyse documents and are known as "bag-of-words" models, since each document is described by a distribution of fixed vocabulary. Using such a representation, methods such as the probabilistic latent semantic analysis (pLSA) [17] and the latent Dirichlet allocation (LDA) [18] can extract coherent topics within document collections in an unsupervised manner.

Fig. 3. Flow chart of the algorithm follows **[19]**.



Some time ago, Fei-Fei and Perona [19] and Sivic et al. [20] applied such methods to the visual domain using [17] and [18] in their algorithm.

They modelled an image as a collection of local patches which were detected by a sliding grid and random sampling of scales. Each patch was represented by a code-word from a large vocabulary of code-words which were sorted in descending order according to the size of their membership and represented simple orientations and illumination patterns. By learning they achieved a model that best represents the distribution of these code-words in each category of scenes. In the recognition process they identified all the code-words in the unknown image. The training and testing process was presented in Fig. 3 in a symbolic way.

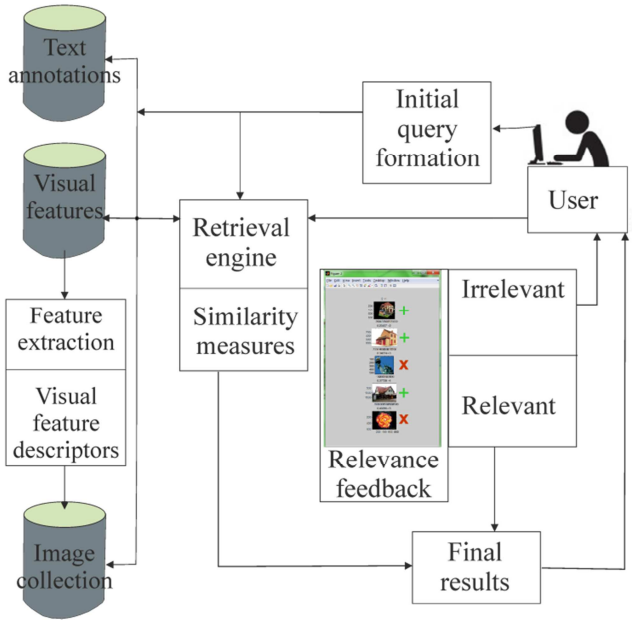
They found the category model that best matched the distribution of the code-words of the particular image. Their model was based on a principled probabilistic approach to learn automatically the distribution of code-words and the intermediate-level themes were treated as texture descriptions.

An advantage of the BoW model is that it is applicable in case of complex indoor and outdoor images. But one of its notorious disadvantages is that the model ignores the spatial relationships among the patches, which are very important in image representation. Additionally, the system needs the preparation of code-words, classes and Bayesian hierarchical models for each class.

D. Relevance Feedback

Large modern DBs actively employ user's interaction for relevance feedback (RF). This is an interactive technique based on feedback information between the user and a search engine in which the user labels semantically similar or dissimilar images with a query image, which is treated as positive and negative samples, respectively. Images labelled in this way are incorporated into a training set. The general architecture of such systems is presented in Fig. 4.

Fig. 4. CBIR architecture with the relevance feedback (RF) mechanism.



A more precisely labelled training set boosts algorithms to build a wider boundary between cluster features. For this purpose, either the Support Vector Machine (SVM) is applied to estimate the density of positive feedbacks or regarding the RF as an two-class only, on-line classification problem or

978-1-4673-8460-5/16/$31.00 ©2016 IEEE

discriminant analysis is used to determine a low dimensional subspace of the feature space, in such a way that positive and negative feedbacks are well separated after being projected onto the subspace.

In recent years, different RF techniques have been suggested to involving the user in the loop to improve the functioning of CBIR [21], [22]. For example, L. Zhang et al [23] introduce a scheme of subspace learning where the training images are associated with only similar and dissimilar pairwise constraints, i.e., Conjunctive Patches Subspace Learning (CPSL) with additional information, to specifically profit the user's previously introduced feedback log data. It means that they minimize the distances between samples with similar pairwise constraints and simultaneously maximize the distances between samples with dissimilar pairwise constraints. Samples are whole images for which neighbourhood is calculated as a locally linear embedding (LLE) [24].

An option of RF is the adaptive technique based on the ostensive model of developing information needs, proposed by J. Urban [25].

Generally, the advantage of the RF approach is the fact that the system can start with a limited number of samples because the user will next provide labelled samples. This approach has enhanced image retrieval accuracy in an efficient way. The disadvantage is that most ongoing systems require several iterations before it receives a stable level, and consequently users lose their patience and may drop it after two or three trials.

E. Semantic Template

In [26] Chang et al. first linked low-level image features with high-level ideas for video retrieval through the semantic visual template (SVT). A visual template is a set of icons or example scenes, or objects belonging to personalized images, such as a crowd, beaches, etc. whose feature vectors are extracted for the query process. In order to build an SVT, the user first determines the template for a specific concept by specifying the objects and their spatial and temporal constraints, the weights assigned to each feature of each object. This initial query outline is put to the system. Through the interaction with users, the system converges to a small list of exemplar queries which are the most relevant (e. g. maximize the recall) the concept in the user's mind.

Firstly, the user selects an annotated image as a query example and adds their concept. Then the system finds visual feature vectors and their weights. According to user hints, the system updates these weights. Having found vector centroids, the ST is received and can be defined as triple $ST = \{C,F,W\}$, where C is the user's concept, F - the feature vector and W - the weight of feature vectors [27].

A disadvantage of this system is the necessity of possessing two databases: an annotating image DB and a big lexical DB [28].

F. WWW Image Retrieval

WWW search engines exploit the evidence from both orthogonal sets of features: the HTML text and the visual, and applied them to two classifiers to recognize a large set of unlabelled images. The URL of an image file often contains a plain hierarchical structure, including some image information, for instance, category of an image. In addition, the HTML document also contains some useful details in the image title, the ALT-tag, the descriptive text surrounding the image, hyperlinks, etc.

However, the disadvantage is the fact that the retrieval precision is poor and as a result the user has to look through the full list to search the required images. This is a time-consuming procedure which always contains multiple combined topics. To boost the Web image retrieval performance, researchers are making an effort to fuse the evidence from textual description and visual image information.

For example, Rasiwasia at al. proposed a combination of a query-by-visual-example (QBVE) with a query-by-semantic-example (QBSE) based on the probability of existence of a visual level represented as a set of feature vectors and the probability of a semantic concept by which an image is annotated. By using the Bayes rule and a similarity function based on methods measuring the distance between two probability distributions (such as the Kullback-Leibler Divergence, Jensen-Shannon Divergence, correlation, etc), they retrieve images most similar to the semantic signature [29].

On the other hand Wang et al. combine the visual features of images with the signatures received from the visual semantic space. For each relevant keyword, a semantic signature of the image is extracted by computing the visual similarities between the image and the reference classes of the keyword using the earlier trained classifiers. The reference classes form the basis of the semantic space of the keyword. If an image has N relevant keywords, then it has N semantic signatures to be computed and stored offline [30].

An advantage of the Web image retrieval is the fact that some extra descriptions on the Web enable the search engine to effectively retrieve semantic-based image information, whereas, the disadvantage is the necessity of having to annotate images in a DB.

G. Our Search Engine with Combined Visual Properties

Our approach is more specific and more user oriented than the above-mentioned approaches. That is why we offer a unique, dedicated user's GUI which allows the user to design their desired image from the image segments. The details of the system are described in [31] and [32].

The system concept is universal. In the construction stage we focus on estate images but for other compound images (containing more than several objects) other sets of classes are needed.

The main concept is presented in Fig. 5. Broadly, our system comprises five principal blocks: the image preprocessing block [33], the classifying unit, the Oracle Database [34], the search engine [35] and the graphical user's interface (GUI). All blocks, except the Oracle DBMS, are implemented in Matlab.

978-1-4673-8460-5/16/$31.00 ©2016 IEEE

A conventional approach to CBIR includes image feature extraction [15], [36]. Our system first segments the new image (e.g. obtained from network resources), generating a set of objects. Each object, selected according to the algorithm is described by some low-level features f_i (see [33]). We select $r = 45$ features for each graphical object, for which we construct a feature vector $\mathbf{O} = \{f_1, f_2, \ldots, f_r\}$.

Subsequently, object classification is prepared based on the feature vector **O**. Objects need to be classified so that they can be used in a spatial object location algorithm and offered to the user as a classified group of objects for semantic selection. To date, the following classifiers have been used in our system:

- decision trees [37], [38];
- a comparison of features of the classified object with a class pattern;
- the Naïve Bayes classifier [39], [40];
- a fuzzy rule-based classifier (FRBC) [41], [32].

The most equivocal objects - those assigned to different classes by the top three classifiers - are identified by the FRBC, which means that the classifier listed last, developed by Ishibuchi in [42] decides which of the three classes a new element belongs to.

Spatial object location has helped reduce the rift between low-level and high-level features in CBIR because, by adding such key information, we can match images more effectively and accurately.

To analyse the spatial layout of objects, a number of methods have been used, for instance: the spatial pyramid representation in a fixed grid [43], the spatial arrangements of regions [44], or the object's spatial orientation relationship [45]. Some researchers have adopted direct image matching, based only on spatial constraints between image regions [46].

In our system, spatial object location is used as the global feature in an image [32]. The objects' mutual spatial relationship is computed on the centroid locations and angles between vectors connecting them, by means of an algorithm designed by Chang and Wu [47] and later adjusted by Guru and Punitha [48], to calculate the first principal component vectors (PCVs).

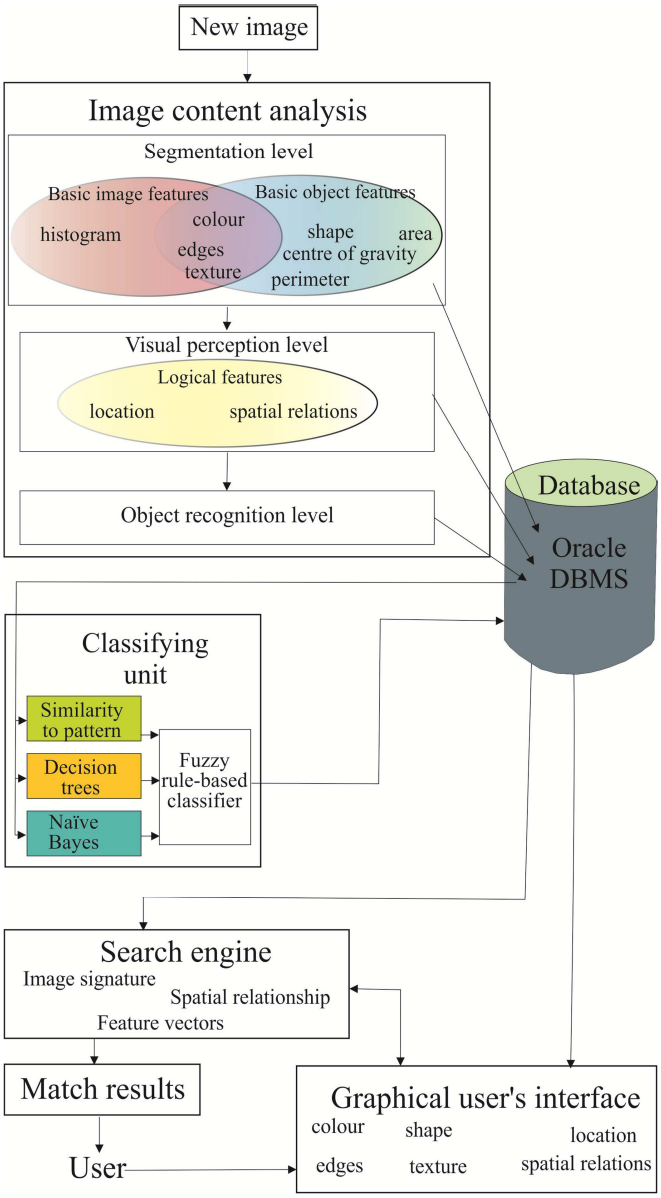
The search engine's modus operandi is reflected by the data architecture and the GUI layout. This GUI has been designed with a view to assisting the user in their attempt to formulate the query they have in mind. First, the user selects a semantic concept by choosing a line sketch and later they design their query. Some of such queries can be really unconventional as we can see in [49].

The following passage examines the similarity between two images, which determines the DB answer to a query. Assuming the query is an image I_q, such as $I_q = \{o_{q1}, o_{q2}, \ldots, o_{qn}\}$, where o_{ij} are objects and the images in the database are I_b, $I_b = \{o_{b1}, o_{b2}, \ldots, o_{bm}\}$, there are M classes of the objects recognized in the database, denoted as labels $L_1, L_2, \ldots, L_M$. In our system we have so far set the number of classes at 40. Then, as the image signature I_i we use the following vector:

$$\mathrm{Signat}(I_i) = [\mathrm{nobc}_{i1}, \mathrm{nobc}_{i2}, \ldots, \mathrm{nobc}_{iM}] \quad (1)$$

where: nobc_{ik} denotes the number of objects o_{ij} of class L_k present in the model of an image I_i.

Fig. 5. The structure of our content-based image retrieval system.



A query image is received from the GUI, where the user designs their own image from selected DB objects. To answer the query I_q, we match it with each image I_b from the database in the first processing step of our search engine. Firstly, we find a similarity measure sim_{sgn} between the signatures of query I_q and image I_b in the following way:

978-1-4673-8460-5/16/$31.00 ©2016 IEEE

$$\mathrm{sim}_{\mathrm{sgn}}(I_q, I_b) = \sum_i (\mathrm{nob}_{qi} - \mathrm{nob}_{bi}) \qquad (2)$$

Eq. (2) is an analogy to the Hamming distance between two vectors of their signatures (cf. (1)), such that $\mathrm{sim}_{\mathrm{sgn}} \geq 0$ and $\max_i(\mathrm{nob}_{qi} - \mathrm{nob}_{bi}) \leq \mathrm{tres}$, tres is the limitation of the number of elements of a particular class by which I_q and I_b can differ. Images with the same classes as the query are preferred. Similarity (2) is asymmetric because we made a strong assumption that images selected from the DB need to have the same classes as the query and that is why the components of (2) can be negative.

If the maximum component of (2) is bigger than a given, as a parameter of the search engine, threshold, then image I_b is discarded. Contrarily, we go to the next stage and we search the spatial similarity $\mathrm{sim}_{\mathrm{PCV}}$ (3) of images I_q and I_b, based on the Euclidean, City block or Mahalanobis metric between their PCVs as:

$$\mathrm{sim}_{\mathrm{PCV}}(I_q, I_b) = 1 - \sqrt{\sum_{i=1}^{3} (PCV_{bi} - PCV_{qi})^2} \qquad (3)$$

If the similarity (3) is smaller than the threshold then image I_b is omitted. The order of steps 2 and 3 is reversible because they are the global parameters and hence can be selected by the user.

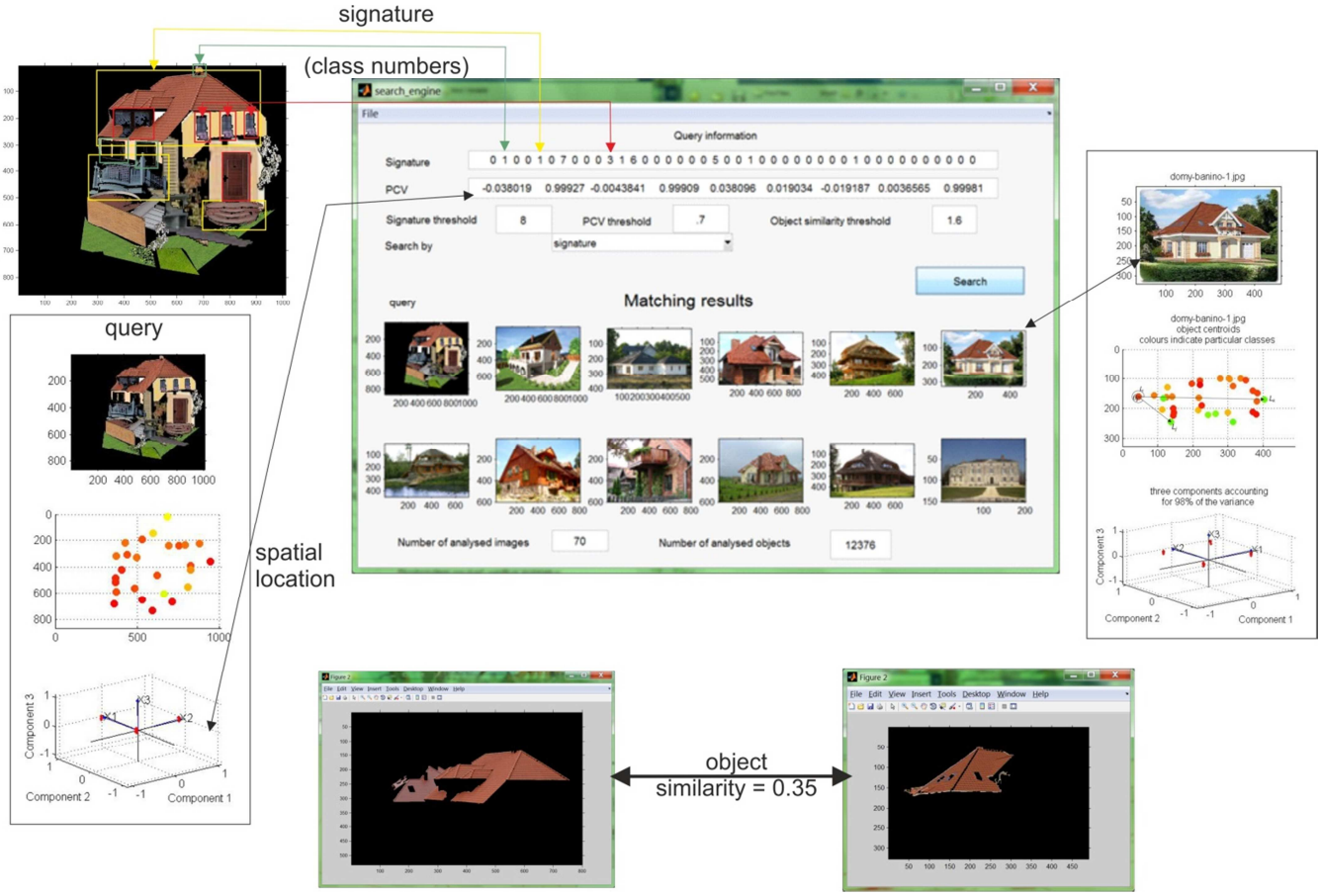
Next, we move to the final stage, that is, we find the similarity of the objects composing both images I_q and I_b. For all objects o_{qi} representing the query I_q, we look for the most similar object o_{bj} of the same class, i.e. $L_{qi} = L_{bj}$. In the case of a lack of object o_{bj} of the class L_{qi}, then $\mathrm{sim}_{\mathrm{ob}}(o_{qi}, o_{bj}) = 0$. Then, similarity $\mathrm{sim}_{ob}(o_{qi}, o_{bj})$ between objects of the same class is calculated based on the Euclidean distance:

$$\mathrm{sim}_{\mathrm{ob}}(o_{qi}, o_{bj}) = 1 - \sqrt{\sum_l (Fo_{qil} - Fo_{bjl})^2} \qquad (4)$$

where l is the index of feature vectors F_O used to represent an object. Hence, we receive a vector of similarities between query I_q and image I_b.

$$\mathrm{sim}(I_q, I_b) = \begin{bmatrix} \mathrm{sim}_{\mathrm{ob}}(o_{q1}, o_{b1}) \\ \vdots \\ \mathrm{sim}_{\mathrm{ob}}(o_{qn}, o_{bn}) \end{bmatrix} \qquad (5)$$

Fig. 6 The main search engine concept.



978-1-4673-8460-5/16/$31.00 ©2016 IEEE

where n is the number of objects that composes the image I_q. In order to compare images I_b with the query I_q, we sum the similarities sim_{ob} (o_{qi}, o_{bj}) and we apply the decreasing order. Therefore, the first some images I_b which obtained the top rank on the similarity list are presented to the user.

Fig. 6 shows the key components of the search engine interface with example images which are contained in the CBIR system. The central window displays the query signature and PCV. Underneath, there are the edit fields to put in threshold values for the signature, PCV and object similarity. At this stage of system verification it is useful to have these thresholds and metrics at hand. In the final Internet version these parameters will be invisible to the user, or limited to the best ranges. The bottom section of the window presents matching results. In the top left of the illustration there is a user designed query consisting of components whose numbers are listed in the signature line. Beneath the query is situated a frame containing a query miniature, a diagram showing the centroids of query elements and, a 3D plot with PCV components. At the bottom of the illustration one can see two elements of the same class (e.g. a roof) whose similarity is calculated. To the right a frame has been placed as an instance of a PCA plot for an image from our DB. The user sets thresholds to establish the type of similarity.

The strong side of our system is its semantic context which limits the semantic gap by taking into account middle-level features, such as objects, their numbers and spatial locations in an image. Additionally, we offer the user the GUI to compose their query by which we eliminate the necessity of looking for a QBE.

On the other hand, our system requires the preparation of a DB containing objects, patterns, and classes

IV. Comparison Results

Most of the currently designed CBIR algorithms use standard databases dedicated especially to image retrieval. They are annotated with class labels to facilitate algorithm testing. The best known collections are:

- the Corel image dataset [50] containing 10,800 images from the Corel Photo Gallery consists of 80 pre-classified concept groups, ranging from sports and houses to outside scenes.
- LA resource pictures [51];
- The Kodak database of true colour images from outside scenes to portraits [52];
- Brodatz textures [53], [54] are examples of mono-chromatic and colour images of textures used to texture feature recognize.
- Images offered by Google used as an additional data source, especially for systems aiming at Web image retrieval [29] [30].
- The *Pascal Visual Object Classes* (VOC) consist of a publicly available dataset of images together with ground truth annotation and standardised evaluation software [55], [56].
- The *ImageNet* is an image database organized according to the WordNet hierarchy in which each node of the hierarchy is depicted by 14,197,122 labelled, high-resolution images, organized by 21841 indexes and belonging to roughly 22, 000 categories. Currently, we have an average of over five hundred images per node [57], [58].
- The *Caltech-256 Image Set* is an image database released in 2006 and consisting of 257 categories of images. It contains 30,608 pictures in total, with 80 to 824 homogeneous pictures per category [59], [60].
- The *Oxford Buildings Dataset* consists of 5062 high resolution (1024×768) images of particular Oxford landmarks [61], [62].

A. User Designed Query

We decided to prepare our own DB for two reasons: (i) when the research began (in 2005) there were few DBs containing buildings which were then at the centre of our attention and (ii) some existing benchmarking databases offered separate objects (like the Corel DB) which were insufficient for our complex search engine concept. At present, our DB contains more than 10 000 classified objects

As we have mentioned, a query is built with the UDQ interface and its number of elements (patches), size and complication depends on the user. Although the user composed only some main details, the search results are quite acceptable (see Table I). For the optimal assigned thresholds a maximum of 11 best matched images from our DB are presented by the search engine.

Table I. The Matching Results for Queries (in the First Row) and the Universal Image Similarity Index for these Results where PCV Similarity is Computed Based on: (Column 1) the Euclidean Distance, (Column 2) the City Block Distance (where: Signature = 17, PCV = 3.5, Object = 0.9), (Column 3) the City Block Distance (where: Signature = 20, PCV = 4, Object = 0.9).

978-1-4673-8460-5/16/$31.00 ©2016 IEEE

B. SIFT and the Google Image Search Engine

In presented context, the comparison of our results with the Google image search engine are also important. The results are presented in TABLE II. We also compare our search engine with the SIFT method and TABLE II column 3 presents the matching results for a query designed in our system. As it can be seen, the best selected matches are those images whose elements can be found in the designed query which is completely consistent with the SIFT assignment.

We have opted for this comparison because these systems match images without annotations, which has been the most important condition. Systems using annotations belong to quite a different category while our focus is on pure image matching.

TABLE II. MATCHES FOR THE GOOGLE AND SIFT IMAGE SEARCH ENGINE (QUERIES IN THE SECOND ROW WITHOUT ANNOTATIONS.)

The Google search engine	The Google search engine	The SIFT search engine

978-1-4673-8460-5/16/$31.00 ©2016 IEEE

C. Discussion

As we can see in Table II the Google engine treats the sketch houses as drawings, not as real photographs, whereas the SIFT one found the images from which the designed query consists, which is proper for this method, but has not been the user's intention who wants to receive house images most similar to their query in general and in detail.

The default comparison of search engines should be carried out based on the standard DB benchmarks. In such a situation, we could find the recall and precision or the SSIM (universal similarity index) [63]. However, in such a way we can only compare if the low-level image features are similar. Whereas, we are aware that the user needs concern more on semantic similarities and in our experiments we shall prepare dedicated search engines for these requirements. Nevertheless, there is no objective mechanism to compare images semantically. That is why we subject images to a qualitative, rather than quantitative, evaluation.

V. CONCLUSIONS

The obtained results seem to be inspiring enough to further elaborate the other stages of the CBIR system, such as the GUI and the search engine. The methods already developed will be also tested with a large number of new classes added to the system. The GUI will also be extended by implementing subclasses to the most general classes.

The optimal parameters for the search engine are being sought in a series of experiments, although, the results we have already achieved, applying the simplest configuration, are rather optimistic.

For future work, we plan to implement an optimised procedure to verify the feasibility of our approach. Additionally, we expect a reasonable performance from the evaluation strategy outlined in the paper.

REFERENCES

[1] Wikipedia, "List of CBIR engines," 2015. [Online]. Available: http://en.wikipedia.org/wiki/List_of_CBIR_engines.

[2] F. Long, H. Zhang and D. D. Feng, "Fundamentals of content-based image retrieval," in *Multimedia Information Retrieval and Management Technological Fundamentals and Applications.*, New York, Sprainger-Verlag, 2003, pp. 1-26.

[3] Y. Liu, D. Zhang, G. Lu and W.-Y. Ma, "A survey of content-based image retrieval with high-level semantics," *Pattern Recognition*, vol. 40, pp. 262-282, 2007.

[4] S. K. Candan and W.-S. Li, "On Similarity Measures for Multimedia Database Applications," *Knowledge and Information Systems*, vol. 3, pp. 30-51, 2001.

[5] J. C. Cubero, N. Marín, J. M. Medina, E. Pons and A. M. Vila, "Fuzzy Object Management in an Object-Relational Framework," in *Proceedings of the 10th International Conference IPMU*, Perugia, Italy, 4-9 July, 2004.

[6] F. Berzal, J. C. Cubero, J. Kacprzyk, N. Marín, A. M. Vila and S. Zadrożny, "A General Framework for Computing with Words in Object-Oriented Programming.," in *International Journal of Uncertainty. Fuzziness and Knowledge-Based Systems.*, vol. 15 (Suppl), Singapore, World Scientific Publishing Company, 2007, pp. 111 131,.

[7] V. Mezaris, I. Kompatsiaris and M. G. Strintzis, "An ontology approach to object-based image retrieval," in *Proceedings of International Conference on Image Processing ICIP 2003.*, 2003.

[8] A. D. Gudewar and L. R. Ragha, "Ontology to Improve CBIR System," *International Journal of Computer Applications*, vol. 52, no. 21, pp. 23-30, 2012.

[9] C. Doulaverakis, E. Nidelkou, A. Gounaris and Y. Kompatsiaris, "A Hybrid Ontology and Content-Based Search Engine For Multimedia Retrieval," in *Workshop Proceedings in Advances in Databases and Information Systems ADBIS '2006*, Thessaloniki, 2006.

[10] O. Allani, N. Mellouli, H. B. Zghal, H. Akdag and H. B. Ghzala, "A Relevant Visual Feature Selection Approach for Image Retrieval," in *VISAPP 2015 - International Conference on Computer Vision Theory and Applications*, Berlin, 2015.

[11] O. Russakovsky and L. Fei-Fei, "Attribute Learning in Large-scale Datasets," in *Proceedings of the 12th European Conference of Computer Vision (ECCV), 1st International Workshop on Parts and Attributes.*, Crete, Greece, 2010.

[12] D. G. Lowe, "Object Recognition from local scale-invariant features," in *International Conferences on Computer Vision*, Corfu, Greece, 1999.

[13] D. G. Lowe, "Distinctive Image Features froam Scale-Invariant Keypoints," *Internationa Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, 2004.

[14] R. Arandjelović and A. Zisserman, "Three things everyone should know to improve object retrieval," in *IEEE Conference on Computer Vision and Pattern Recognition*, Providence, RI, USA, 2012.

[15] T. Tuytelaars and K. Mikolajczyk, "Local Invariant Feature Detectors: A Survey," *Computer Graphics and Vision*, vol. 3, no. 3, p. 177–280, 2007.

[16] E. Rublee, V. Rabaud , K. Konolige and G. Bradski, "ORB: an efficient alternative to SIFT or SURF," in *IEEE International Conference on Computer Vision (ICCV)*, Barcelona, Spane, 6-12, Nov, 2011.

[17] T. Hofmann, "Probabilistic latent semantic analysis," in *Proceedings of the15th Conference on Uncertainty in Artificial Intelligence*, Stockholm, 1999.

[18] D. M. Blei, A. Y. Ng and M. I. Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993-1022, 2003.

[19] L. Fei-Fei and P. Perona, "A Bayesian Heirarcical Model for Learning Natural Scene Categories," in *Computer Vision & Pattern Recognition CVPR*, 2005.

[20] J. Sivic, B. C. Russell, A. A. Efros, A. Zisserman and W. T. Freeman, "Discovering objects and their location in images," in *Proceedings of Internationa Conference of Computer Vision*, Beijing, 2005.

[21] L. Zhang, L. Wang and W. Lin, "Generalized biased discriminant analysis for content-based image retrieval," *IEEE Transactions on System, Man, Cybernetics, Part B - Cybernetics*, vol. 42, no. 1, pp. 282-290, 2012.

[22] L. Zhang, L. Wang and W. Lin, "Semi-supervised biased maximum margin analysis for interactive image retrieval," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 2294-2308, 2012.

[23] L. Zhang, L. Wang and W. Lin, "Conjunctive patches subspace learning with side information for collaborative image retrieval," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3707-3720, 2012.

[24] S. T. Roweis and L. K. Saul, "Nonlinear Dimensionality Reduction by Locally Linear Embedding," *Science*, vol. 290, no. 5500, pp. 2323-2326, Dec. 2000.

[25] J. Urban, J. M. Jose and C. J. van Rijsbergen, "An adaptive technique for content-based image retrieval," *Multimedial Tools Applied*, no. 31, pp. 1-28, July 2006.

[26] S.-F. Chang, W. Chen and H. Sundaram, "Semantic Visual Templates: Linking Visual Features to Semantics," in *International Conference on*

978-1-4673-8460-5/16/$31.00 ©2016 IEEE

Image Processing, 1998. ICIP 98., Chicago, 1998.

[27] Y. Zhuang, X. Liu and Y. Pan, "Apply Semantic Template to Support Content-based Image Retrieval," in *the Proceeding of IS&T and SPIE Storage and Retrieval for Media Databases 2000*, San Jose, California, USA, Jan, 2000.

[28] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross and K. Miller, "Introduction to WordNet: An On-line Lexical Database," *Communications of the ACM*, vol. 38, no. 11, pp. 39-41 , Nov. 1995.

[29] N. Rasiwasia, P. J. Moreno and N. Vasconcelos, "Bridging the Gap: Query by Semantic Example," *IEEE TRANSACTIONS ON MULTIMEDIA*, vol. 9, no. 5, pp. 923-938, Aug 2007.

[30] X. Wang, S. Qiu, K. Liu and X. Tang, "Web Image Re-Ranking Using Query-Specific Semantic Signatures," *IEEE Transactions on Pattern Analysis and Machine Intelligence* , vol. 36, no. 4, pp. 810-823, April 2014.

[31] T. Jaworska, "Multi-criteria object indexing and graphical user query as an aspect of content-based image retrieval system.," in *Information Systems Architecture and Technology*, L. Borzemski, A. Grzech, J. Świątek and Z. Wilimowska, Eds., Wrocław, Wrocław Technical University Publisher, 2009, pp. 103-112.

[32] T. Jaworska, "Spatial representation of object location for image matching in CBIR," in *New Research in Multimedia and Internet Systems*, vol. 314, A. Zgrzywa, K. Choroś and A. Siemiński, Eds., Wrocław, Springer, 2014, pp. 25-34.

[33] T. Jaworska, "Object extraction as a basic process for content-based image retrieval (CBIR) system," *Opto-Electronics Review*, vol. 15, no. 4, pp. 184-195, December 2007.

[34] T. Jaworska, "Database as a Crucial Element for CBIR Systems," in *Proceedings of the 2nd International Symposium on Test Automation and Instrumentation*, Beijing, China, 16-20 Nov., 2008.

[35] T. Jaworska, "A Search-Engine Concept Based on Multi-Feature Vectors and Spatial Relationship," in *Flexible Query Answering Systems*, vol. 7022, H. Christiansen, G. De Tré, A. Yazici, S. Zadrożny and H. L. Larsen, Eds., Ghent, Springer, 2011, pp. 137-148.

[36] Y.-J. Zhang, Y. Gao and Y. Luo, "Object-Based Techniques for Image Retrieval," in *Multimedia Systems and Content-Based Image Retrieval*, S. Deb, Ed., Hershey, London, IDEA Group Publishing, 2004, pp. 156-181.

[37] U. M. Fayyad and K. B. Irani, "The attribute selection problem in decision tree generation," in *the 10th National Conference on Artificial Intelligence, AAAI*, 1992.

[38] J. R. Quinlan, "Induction of Decision Trees," *Machine Learning*, vol. 1, pp. 81-106, 1986.

[39] C. Faloutsos, R. Barber, M. Flickner, J. Hafner, W. Niblack and D. Petkovic, "Efficient and Effective Querying by Image Content.," *Journal of Intelligent Information Systems*, vol. 3, pp. 231-262, 1994.

[40] I. Rish, "An empirical study of the naive Bayes classifier," in *IJCAI-2001 workshop on Empirical Methods in AI*, 2001.

[41] H. Ishibuchi and Y. Nojima, "Toward Quantitative Definition of Explanation Ability of Fuzzy Rule-Based Classifiers," in *IEEE International Conference on Fuzzy Systems*, Taipei, Taiwan, June 27-39, 2011.

[42] T. Jaworska, "Application of Fuzzy Rule-Based Classifier to CBIR in comparison with other classifiers," in *11th International Conference on Fuzzy Systems and Knowledge Discovery*, Xiamen, China, 19-21.08.2014.

[43] G. Sharma and F. Jurie, "Learning discriminative spatial representation for image classification," in *Proceedings of the British Machine Vision Conference 2011*, Dundee, 2011.

[44] J. R. Smith and S.-F. Chang, "Integrated spatial and feature image query," *Multimedia Systems*, no. 7, p. 129–140, 1999.

[45] X. M. Zhou, C. H. Ang and T. W. Ling, "Image retrieval based on object's orientation spatial relationship," *Pattern Recognition Letters*, vol. 22, pp. 469-477, 2001.

[46] T. Wang, Y. Rui and J.-G. Sun, "Constraint Based Region Matching for Image Retrieval," *International Journal of Computer Vision*, vol. 56, no. 1/2, pp. 37-45, 2004.

[47] C.-C. Chang and T.-C. Wu, "An exact match retrieval scheme based upon principal component analysis," *Pattern Recognition Letters*, vol. 16, pp. 465-470, 1995.

[48] D. S. Guru and P. Punitha, "An invariant scheme for exact match retrieval of symbolic images based upon principal component analysis," *Pattern Recognition Letters* , vol. 25, p. 73–86, 2004.

[49] J. Deng, J. Krause, A. Berg and L. Fei-Fei, "Hedging Your Bets: Optimizing Accuracy-Specificity Trade-offs in Large Scale Visual Recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Providence, RI, USA, 2012.

[50] Corel comp., "The COREL Database for Content based Image Retrieval".

[51] Z. Yang and C.-C. Jay Kuo, "Learning image similarities and categories from content analysys and relebance feedback," in *Proceedings of the ACM Multimedia Workshops. Multimedia00'*, Los Angeles, CA, USA , Oct 30 - Nov 03, 2000.

[52] the Eastman Kodak Company, [Online]. Available: http://r0k.us/graphics/kodak/.

[53] D.-C. He and A. Safia, "Multiband Texture Database," 2015. [Online]. Available: http://multibandtexture.recherche.usherbrooke.ca/.

[54] D.-C. He and A. Safia, "New Brodatz-based Image Databases for Grayscale Color and Multiband Texture Analysis," *ISRN Machine Vision*, pp. 1-14, Article ID 876386, 2013.

[55] M. Everingham, A. S. Eslami, L. Van Gool, C. K. I. Williams, J. Winn and A. Zisserman, "The PASCAL Visual Object Classes Challenge: A Retrospective," *International Journal of Computer Vision*, no. 111, p. 98–136, 2015.

[56] M. Everingham, L. Van Gool, C. K. I. Williams, A. Zisserman, J. Winn, A. S. Eslami and Y. Aytar , "The PASCAL Visual Object Classes Homepage," 2015. [Online]. Available: http://host.robots.ox.ac.uk/pascal/VOC/index.html.

[57] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in *IEEE Conference on Computer Vision and Pattern Recognition*, Miami, USA, June, 2009.

[58] L. Fei-Fei, K. Li, O. Russakovsky, J. Krause, J. Deng and A. Berg, "ImageNet," Stanford Vision Lab, Stanford University, Princeton University, 2014. [Online]. Available: http://www.image-net.org/.

[59] G. Griffin, A. D. Holub and P. Perona, "The Caltech 256," California Institute of Technology, Los Angeles, 2006.

[60] G. Griffin, "Caltech256," 2006. [Online]. Available: http://www.vision.caltech.edu/Image_Datasets/Caltech256/.

[61] J. Philbin, O. Chum and M. a. S. J. a. Z. A. Isard, "Object Retrieval with Large Vocabularies and Fast Spatial Matching," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2007.

[62] J. Philbin, R. Arandjelović and A. Zisserman, "The Oxford Buildings Dataset," Department of Engineering Science, University of Oxford, Nov 2012. [Online]. Available: http://www.robots.ox.ac.uk/~vgg/data/oxbuildings/.

[63] Z. Wang , A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image Qualifty Assessment: From Error Visibility to Structural Similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, p. 600–612, April 2004.

978-1-4673-8460-5/16/$31.00 ©2016 IEEE