2014

Content-based image retrieval over IEEE 802.11b noisy wireless networks

Chatzistavros, Evaggelos

Wiley

INTERNATIONAL JOURNAL OF COMMUNICATION SYSTEMS
Int. J. Commun. Syst. 0000; **00**:1–21
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/dac

Content-Based Image Retrieval over IEEE 802.11b Noisy Wireless Networks

Evaggelos Chatzistavros*, Savvas A. Chatzichristofis, Konstatinos Zagoris, and George Stamatelos

Department of Electrical & Computer Engineering, Democritus University of Thrace, 67100, Xanthi, Greece

SUMMARY

Mobile devices such as smartphones and tablets are widely used in everyday life to perform a variety of operations, such as e-mail exchange, connection to social media, bank/financial transactions etc. Moreover, due to the large growth of multimedia applications, video and image transferring and sharing via a wireless network is becoming increasingly popular. Several modern mobile applications perform information retrieval and image recognition. For example, Google Goggles is an image recognition application which is used for searches based on pictures taken by handheld devices. In most of the cases, image recognition procedure is an image retrieval procedure. The captured images or a low level description of them are uploaded online, and the system recognizes their content by retrieving visually similar pictures. Taking into account the last comment, our goal in this paper is to evaluate the process of image retrieval/ recognition over an IEEE 802.11b network, operating at 2.4 GHz. Our evaluation is performed through a simulated network configuration, which consists of a number of mobile nodes communicating with an Access Point (AP). Throughout our simulations we examine the impact of several factors, such as the existence of a strong Line of Sight (LOS) during the communication between wireless devices. Strong LOS depends on the fading model used for the simulations and has an effect on Bit Error Rate (BER). We have used a large number of image descriptors and a variety of scenarios, reported in the relative literature, in order to comprehensively evaluate our system. To reinforce our results, experiments were conducted on two well known images databases using ten descriptors from the literature.

Received . . .

KEY WORDS: IEEE 802.11b, Image Retrieval, Bit Error Rate, Multimedia Applications, Noisy Environment

1. INTRODUCTION

Connection to a network while being in a public area is one of the most common, modern and advancing ways of communication. There is a great variety of mobile devices, such as cell phones, smartphones, pdas, laptops, netbooks etc., which have the capability of connecting to a network,

*Correspondence to: E-mail: ec3526@ee.duth.gr

either 3G or wireless. Moreover, there are lots of applications, many of which perform information retrieval and image recognition, such as Google Goggles † and Nokia Point and Find ‡, which allow a user to perform image retrieval based on an photo taken by their mobile device. Such applications are used by the end user for the identification of a variety of labels, book covers, landmarks, etc. E.g. on an outdoor scenario, a user can take a photo of, let's say, the Parthenon, and using Google Goggles the photo is uploaded to a server where a search procedure takes place. After a while, information about the monument such as history, the architecture and its function, is downloaded to the user's mobile device.

However, the ease of connection to a wireless network comes with a price: in many cases the large number of available networks results in a large number of different transmitting signals. Moreover, the topography of an area, and/ or the surrounding buildings in the case of an urban area, has an effect on signal quality, due to signal fading. As reported in [1], the aforementioned problems cause an increase in the Bit Error Rate (BER) of a specific wireless channel, thus affecting the quality of the transmitted information. In such a "hostile" environment, it is vital for a user to be able to use an application without having to deal with inaccuracies or doubts about the outcome of the particular task. A comparison of techniques for reducing BER in an indoor scenario can be found in [2].

In this paper, we will examine the case of transmitting data over an erroneous channel using the UDP (User Datagram Protocol) [3]. A user performing image retrieval using an application on his/ hers mobile device, demands real-time response in terms of the results. Due to the protocol's simple implementation which introduces minimum delay and zero retransmissions, UDP is appropriate to be used for such an application, as the identification of a statue or painting in a museum followed/ complemented by the appropriate information or literature. Moreover, the optional use of the checksum field makes it the ideal candidate for an application similar to the scenarios discussed in this paper. More details on the UDP protocol are presented in Section 3. Our goal is to evaluate a large number of image retrieval methods, through simulations on wireless networks, presenting the corresponding results. The aforementioned evaluation of image retrieval methods over noisy channels, takes into account fading conditions, surrounding environment etc. We emphasize on the fading model selected for our simulations, since it directly affects channel quality, defines channel conditions, noise, LOS, BER etc. Although we cannot examine every possible network configuration, we are confident that our work is a first step towards this direction, since we take into account a serious amount of network parameters. The simulated scenario represents a well-defined, realistic configuration, very similar to the network topologies of every day network communications.

Apart from evaluating a variety of different image retrieval methods, we also evaluate the way image retrieval is performed. To be more specific, we consider two different cases: in the first case, the image descriptor is transmitted through the network, while in the second case the image is transmitted through the network according to [4]. Our goal is to examine whether or not it is more effective to extract the appropriate descriptor at the user's side, and then transmit it to the server over the noisy channel, than transmitting the image from the user to the server where the image retrieval process takes place. Through this process we will be able to evaluate and propose one of the existing/ examined methods of image retrieval, which has satisfying performance under the

† http://www.google.com/mobile/goggles/
‡ http://pointandfind.nokia.com

DOI: 10.1002/dac

simulated circumstances. We must mention that we examine only the wireless part of a client - server transmission, since the number of errors over a wired link is in fact very low.

Figure 1 illustrates an artistic representation of an indoor image retrieval scenario, e.g. in a museum. During a tour in a museum, a woman finds a painting very interesting and wants to obtain more information about it, so she takes a photo using her mobile device. A similar scenario would be during a biology student's walk in a botanic garden. In this case, the student could retrieve all necessary information related to a plant species with just one click on his mobile device. A retrieval system could also be employed for localization purposes. Someone walking in a building complex, i.e. a large hospital, could find his/hers exact location in the building by taking a photo of the surroundings of his/hers current position. In the first two examples, a user performs object recognition, while in the third one he performs scene recognition. Irrespectively of the differences between the aforementioned examples, there are two cases regarding the image retrieval process: either the image descriptor is extracted locally on the mobile device and then sent through the network, or the image itself. The process takes place in real time and the data file (descriptor or image) is uploaded to a server by the UDP protocol, where the actual image retrieval process takes place. Finally, the corresponding data are downloaded to the user's mobile device.

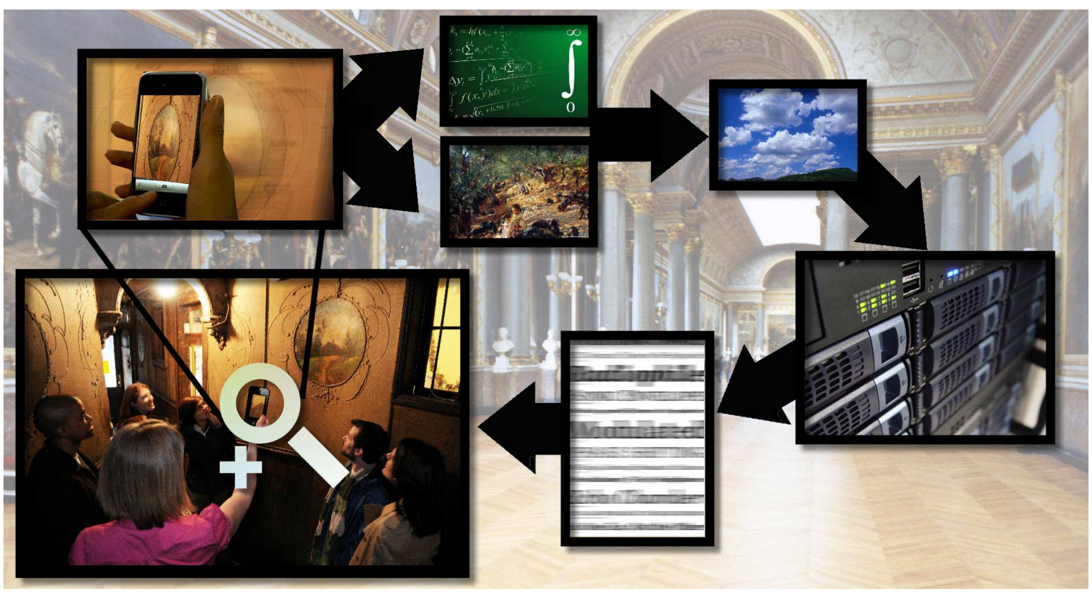

Figure 1. An Artistic Representation of the Indoor Image Retrieval Scenario.

The structure of this paper is as follows: in Section 2, we present relevant work performed on image retrieval over wireless networks. Section 3 describes the configuration and implementation of our simulated scenario, the two possible ways to perform image retrieval process and the fading model applied in our simulations. In the sequel, Section 4 presents the results obtained from our experiments and finally, our conclusions are drawn in Section 5.

Prepared using dacauth.cls
DOI: 10.1002/dac

2. RELATED WORK

In previous work, several authors examine content-based image retrieval for applications used on mobile devices, such as smartphones, or conduct comparative performance evaluation of different image retrieval methods. However, none has conducted, to our knowledge, an evaluation of several image retrieval methods under a hostile environment, such as a noisy wireless channel.

Previous work in image retrieval over wireless networks is reported in [4–9]. A review on current advances in the field of content-based image retrieval is presented in [4]. Moreover, authors analytically present all individual steps of the image retrieval process, such as feature extraction, feature indexing and matching, and geometric verification. An appropriate system able to support this process is also presented and evaluated in terms of many parameters, such as retrieval accuracy, latency, energy consumption of the mobile device and network type (3G, WLAN). Authors in [5] present a client-server content-based image retrieval framework, designed for mobile platforms. A photo taken from the mobile device, the client, is transmitted to a server which performs content-based image retrieval from a database and then sends the query results back to the client over a wireless network connection. A more specialized application is developed in [6], where authors propose a scheme of performing garment image retrieval on the WEB over a wireless network. A similar work is performed in [7], where the authors present a client-server system using the Compressed Histogram of Gradients (CHoG) descriptor to perform image retrieval using a 3G network, also proposing a method of reducing overall latency of the system. In [8] authors evaluate the performance of a variety of descriptors, MPEG-7 image signatures, Compressed Histogram of Gradients (CHoG) and Scale Invariant Feature Transform (SIFT). Their goal is to examine which descriptor among the three is appropriate to be used in mobile search applications. A system to be used on mobile phones, named Mobile-based Segment and Image Retrieval (MOSIR) is presented in [9], using edge-based and color-layout features to perform content-based image retrieval.

3. SYSTEM OVERVIEW

Our system consists of two major parts; the wireless network and the image retrieval system. In this section, we present a thorough presentation and analysis of both of them.

3.1. Network Configuration

As previously mentioned, our goal in this paper is to conduct an evaluation of image retrieval process over a wireless network. The network configuration to achieve this goal, is as follows: we simulate an IEEE 802.11b wireless network, operating at 2.4 GHz, using IEEE Distributed Coordination Function CSMA/CA as the multiple access scheme. The aforementioned configuration is the one defined by the IEEE 802.11 standard [10]. Our simulations were conducted using the Network Simulation 2 (NS-2) [11]. We assume a wireless network, distributed over 500 m $\times$ 500 m area, consisting of 10 mobile nodes randomly placed within the area and an Access Point (AP), located in the middle of the area. The selection of the area ensures the existence of a variety of different conditions. First of all, the Ricean fading model [12] presented later in the manuscript, is applied in order to simulate an outdoor, urban area, consisting of buildings, which may affect wireless signal

DOI: 10.1002/dac

quality. In such an area, a node may communicate in both ad-hoc (those being far away from the Access Point) and infrastructure (directly connected to the Access Point) mode. This in turn has an effect to the number of bit errors introduced to a data packet, since more hops mean greater bit error probability. Apart from the outdoor environment, the selection of the specific area and fading model, gives us the possibility to simulate a large indoor space, such as a large mall or a museum, by properly selecting the Ricean model K-factor. More details on Ricean model and the K-factor are given in the next paragraphs. In order to get more accurate results, we used 10 different initial topologies, for each of which we performed 10 different simulations with different (random) seed every time, resulting in a total number of 100 simulations. All nodes have a transmission range of approximately 100 m, thus some distant nodes may not be able to communicate directly with the AP. In such a case, a distant node may communicate with the AP through intermediate nodes within its transmission range, operating in ad hoc mode as previously stated. All nodes use a very popular routing protocol for wireless networks, the Ad hoc On-Demand Distance Vector (AODV) [13] as their routing protocol, and the UDP as the transport protocol.

The reason why we chose UDP as the transport protocol is because we consider it to be more suitable for our configuration compared to TCP (Transmission Control Protocol) [14]. First of all, UDP offers reduced overhead and is much faster compared to TCP. In practice, there must exist an immediate response and interaction between a user and the application, therefore speed is a vital feature since the image retrieval process discussed is a real time application. Moreover, UDP is appropriate for transferring small size packets, and the 'checksum' field is optiona, in addition to TCP [15]. Therefore, it does not support error correction, since it depends on the application whether or not it will detect and/or recover a lost or damaged packet. This is exactly what we want to examine, an application's ability to perform image retrieval when either a descriptor or image file is corrupted by transmission errors. In addition to this, based on IEEE Std 802.11, there is not any mention that one or more error correction mechanisms, such as FEC [16], are implemented in IEEE 802.11b (Clause 17).[§]

A similar platform to the one described above will be implemented and integrated in two Greek river basins. The goal of the system will be to sufficiently transfer data extracted from sensors and digital cameras to a central access point. The system will consist of an Ad Hoc wireless network, using either IEEE 802.11b/g or IEEE 802.15.4 standard, and data will be forwarded to an access point. Our goal is to perform data mining based on the collected data not allowing channel errors to affect the collected data quality.

Another very important parameter in network simulations and in our configuration specifically is the choice of the fading model. A simulation is the imitation of a real world process, thus a number of mathematical fading models have been developed/proposed to meet certain needs depending on the configuration we want to study. Freespace, Two-Ray Ground, Rayleigh and Rician are four of the most popular fading models used in simulations [17].

Rician fading model is a mathematical model used to simulate a wireless link and is regarded as the most realistic fading model among the others. Moreover depending on the value of a parameter called the K factor, we can simulate all other models. The K-factor is equal to the ratio between

[§]Notice that in [10], the FEC mechanism is taken into account regarding IEEE 802.11a and IEEE 802.11n (Clause 18 and Clause 20 respectively)

DOI: 10.1002/dac

direct-path power (Line-of-Sight (LOS) power) and diffuse power, and is expressed linearly. The greater the K-factor value, the stronger the direct-path power, thus two communicating nodes have a much stronger LOS. In case where $K = 0$, there is no LOS therefore a Rayleigh fading channel is simulated, whereas if $K = 1000$ there is a very strong LOS, thus we simulate an AWGN channel. In [12] the average BER over a Rician channel is calculated as:

$$P_{b\_CCK} = \frac{2^{k-1}}{2^k - 1} \left(\sum_{m=1}^{M-1} \frac{(-1)^{m+1} \begin{pmatrix} M-1 \\ m \end{pmatrix} (1+K)}{1 + K + m(1 + K + \overline{\gamma_s})} \right) \exp\left(-\frac{K m \overline{\gamma_c}}{1 + K + m(1 + K + \overline{\gamma_s})} \right) \quad (1)$$

Where M is the number of bits used for M-ary Bi-Orthogonal keying, $M = 8$ for 11 Mbps, $k = log_2 M$, γ_s=the average symbol SNR, equal to k times of average SNR per bit and γ_c = the average chip SNR, where each chip is a complex QPSK bit-pair, at a chip rate 11Mchips/s.

Figure 2 shows BER values for a Rician fading channel with $K = 20$. Each line represents BER for a specific data rate and velocity, based on the experimental results obtained by the authors.

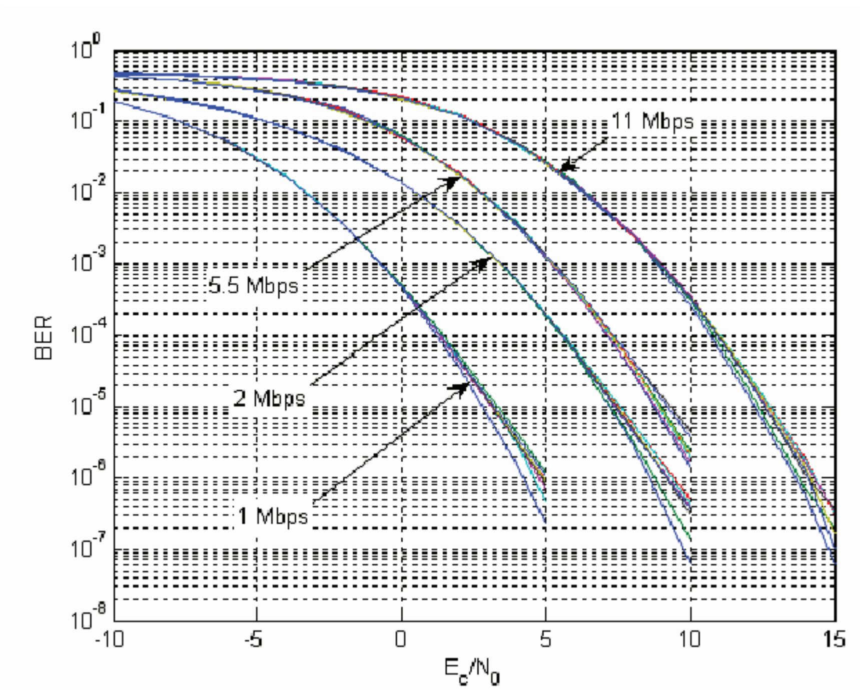

Figure 2. Bit Error Rate for Rician Fading Channel for $K = 20$ [12].

Authors in [18] and [19] have measured the K factor on real environments and their results are very close to the ones proposed by the theoretical model of [12], in terms of the estimated K factor for different environments. We have used Rician fading model with K-factor $K = 20$ in our configurations, since it is the most appropriate for our scenarios.

Prepared using dacauth.cls
DOI: 10.1002/dac

3.2. *Image Retrieval*

Any technology that helps to organize digital images collections based on their visual content is characterized as Content Based Image Retrieval (CBIR) technique. [20]. Therefore, anything ranging from an image similarity method to a complicate image annotation engine is designated under the term CBIR. Generally, CBIR methods can be classified into two categories based on the visual features employed: (i) CBIR based on global features and (ii) CBIR based on local features.

Global features have extensively been used in literature and are able to describe the visual content of an image in a holistic way. The information described by these descriptors concerns either the color, the shape, or the texture of the image. Global descriptors are able to retrieve images with similar visual properties to the query image; e.g. if the dominant color of the query image is red, retrieved images are also red. In applied research, CBIR often relies on global features, at least as a foundation for further research [21]. Additionally, global features have also been adopted in several other multimedia indexing applications such as video summarization applications [22]. On the other hand, CBIR-methods based on Local Features (LFs), aim to describe the visual content of the images in a more semantic way, retrieving images with similar content apart from the visual likeness. The extraction of LFs, indexing and similarity matching takes place in high dimensional feature space, and thus a considerable complexity is introduced. In all cases, development of an ideal CBIR system must be designed to accurately retrieve images from any image database, irrespectively of their size for any given query.

Current research in the corresponding field has seen increased effort to merge the two above categories in order to eliminate problems raised from each one and speed up the retrieval process. The most common approach of such systems is the Bag-Of-Visual-Words (VW) paradigm. It is a promising framework [23] and has been widely adopted mainly because of its two major advantages: (i) better retrieval effectiveness against global features on nearly identical images and (ii) better efficiency against local features [24]. More details about the basics of bag-of-visual-words paradigm and a brief review of the CBIR methods that employ the VW model can be found in [25].

In the case of human perception, the procedure of scene and object recognition is generally effective, automatic and reliable. On the other hand, in case of computers, this procedure is upscaled into an extremely difficult and highly computational problem. Each object in the world can cast an infinite number of different 2D images onto the retina as the object's position, pose, lighting, and background vary relative to the viewer [26]. Many successful approaches for scene or object recognition transform low-level global or local descriptors into richer representations of intermediate complexity [27]. Several image retrieval techniques have been adopted in order to achieve robust object and scene recognition [28], especially those employing the BOVW model [29–31]. This procedure is also known as object-based image retrieval (OBIR) [32]. It is worth noting that, in the past, content-based image retrieval methods have also been adopted in robot grasping systems to facilitate object recognition [33, 34]. OBIR has led to a variety of applications similar to the ones mentioned in Section 1, employed by users over a variety of different situations, such as obtaining information about works of art, monuments, brands etc. Object recognition using traditional image retrieval techniques requires less computational power than other object recognition methods, such as moments etc., thus making such techniques appropriate for mobile

DOI: 10.1002/dac

devices' applications in terms of resource allocation and computational time. Yet, the question arising is '*how to employ CBIR techniques for object and scene recognition*'?

We revisit the museum scenario described earlier in Section 1. Assume that images of all the paintings exhibited in the museum are available, stored together with their low level descriptors, in a web or local server. Additionally, assume that, more than one images from each exhibit, taken under different angle or lighting conditions are also available. These images will form the database in which the searching procedure takes place. Moreover, stored images are accompanied by metadata such as the name of the painting, the artist etc. When a visitor of the museum uploads a query to the server, the distance between the query image and the images in the database is calculated using an appropriate similarity matching technique. In general, the objective of an image retrieval system is to retrieve images in rank order, where the rank is determined from the relevance to the query at hand. In other words, CBIR methods typically rank the whole collection via a distance measure. The images with lowest distances are ranked highest in the retrieval process while the images with highest distances are ranked lowest. In case of object and scene recognition, the user is concerned with the validity of the first results, without requiring the system to retrieve the entire ground truth of the query. Therefore in this case, the system returns to the user the information accompanying the images with the smallest distance from the query one. In other words, when a user uploads an image of a painting, receives the metadata that follow the most visually similar stored images.

All the experiments were performed using img(Rummager) [35] open source image retrieval engine, employing the following global descriptors:

- Color and Edge Directivity Descriptor (CEDD) [36, 37] - *Compact Composite Descriptor*
- Fuzzy Color and Texture Histogram (FCTH)[36, 38] - *Compact Composite Descriptor*
- Edge Histogram Descriptor (EHD) [39] - *MPEG-7 Descriptor*
- Color Layout Descriptor (CLD) [39] - *MPEG-7 Descriptor*
- Spatial Color Distribution Descriptor (SpCD)[40] - *Compact Composite Descriptor*
- Tamura texture features [41]
- RGB Color Histograms [42]

We also evaluate the Bag-Of-Visual-Words (VW) model [43], using three dictionaries of different size, 128, 512 and 2048 visual words. Visual words were generated using SURF [44] features and k-means classifier.

In order to evaluate each descriptor's performance, we used the Mean Average Precision (MAP) metric:

$$\text{Precision} = P = \frac{\text{Number of relevant images retrieved}}{\text{Total number of images retrieved}} \tag{2}$$

$$\text{Recall} = R = \frac{\text{Number of relevant images retrieved}}{\text{Total number of relevant images}} \tag{3}$$

The average precision AP for a single query q is the mean over the precision scores after each retrieved relevant item:

$$AP(q) = \frac{1}{N_R} \sum_{n=1}^{N_R} P_Q(R_n) \tag{4}$$

Prepared using dacauth.cls
DOI: 10.1002/dac

where R_n is the recall after the nth relevant image was retrieved. N_R is the total number of relevant documents for the query. The mean average precision *MAP* is the mean of the average precision scores over all queries:

$$MAP = \frac{1}{|Q|} \sum_{q \in Q} AP(q) \tag{5}$$

where Q is the set of queries q.

A user who wants to employ a web based image retrieval system, such as Google Images¶, is mainly interested on the optimum presentation of the relevant results in the first pages, since a relevant result may not be detected by the user in case it is presented in further pages. Those systems are highly correlated with early results. Precession at 10 (P@10) describes the system's capability to retrieve as many results as possible in the early 10 positions. In the case of Nister database, each ground truth contains only 4 images. For this reason we only use the P@4.

4. EXPERIMENTAL RESULTS

In this section we present experimental results of the image retrieval process using two different scenarios. In the first case the descriptor is extracted to the mobile device and transmitted to the server through the wireless network, while in the other case an image is transmitted through the network to the server. In both cases, image retrieval process is performed on the server. Experiments were conducted on two well known databases: Initially experiments performed using the *UKBench* [45] (also known as 'Nister') database, employing 100 randomly selected queries. The Nister database, consists of 10200 images arranged in 2550 groups of four images each. Each group includes images of a single object. The pictures are taken from different viewpoints and occasionally under different lighting conditions. The first image of every object is used as a query image. Given a query image, only images from the same group are considered relevant. In case of an *indoor scenario*, mobile visual search employs several specific semantic queries. E.g., suppose a user who takes a picture of the cover of a book he is interested in purchasing, using his/hers mobile device. The user can perform content based image retrieval and get information of a number of relevant stores (online or not) where he/she can purchase the book from. The Nister database provides the capability of performing such content based image retrieval based on the fact that it consists of several semantic based images, such as book and/or CD covers etc.

In the sequel, experiments performed on the Wang database [46]. The Wang database is a subset of 1000 manually-selected images from the Corel stock photo database and forms ten classes of 100 images each. In particular, queries and ground truths proposed by the MIRROR image retrieval system [47] are used. MIRROR separates the Wang database into 20 queries. Although the size of the database is small, it is a wide-spread image retrieval evaluation dataset, especially for the evaluation of the effectiveness of the global descriptors. Wang database includes query images where the ground truth consist of pictures with similar visual content to the query image, without

¶http://images.google.com/

DOI: 10.1002/dac

this implying the co-occurrence of the same objects. Experiments using the the Wang database are appropriate for simulating an *outdoor image retrieval* scenario.

4.1. Transmitting the descriptor

In this setup we assume that a descriptor F of a query image Q is extracted locally to the mobile device and is then used by a user A in order to perform image retrieval on a database D stored in a server S, transmitted to the Access Point through the wireless network. A descriptor is a vector which describes the visual content of an image. In this setup the transmitted amount of information through the wireless channel is minimized, since a descriptor contains a significantly lower amount of information compared to an image. It is transmitted as a sequence of bits, thus is affected by errors caused on the wireless channel, altering a number of bits from 0 to 1 and vice versa, depending on BER on the wireless channel.

Figure 3 illustrates the Mean Average Precision variation over different Bit Error Rate values for the experiments performed on the Nister database.

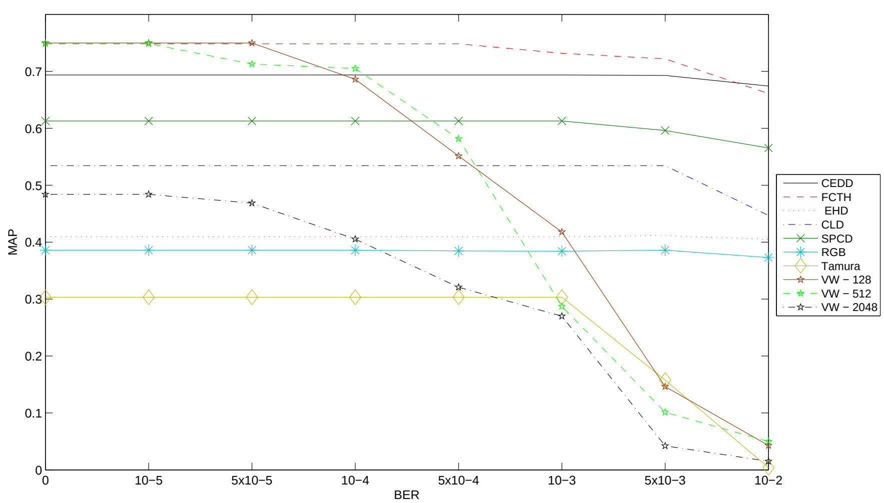

Figure 3. Nister Database:MAP vs BER

As seen from Figure 3, most of the descriptors maintain their initial values even for BER values up to 10^{-3}. Compact composite descriptors (CEDD, FCTH and SpCD) appear to have the highest MAP values among the others, except from Visual Words VW - 512. Moreover, compact composite descriptors tend to maintain their initial scores, with FCTH being the one with the highest MAP value except the case of BER=0.01. CEDD, FCTH and SpCD present an overall degradation of 2.82 %, 11.66 % and 16.36 % respectively. Visual Words, regardless of the dictionary size used, do not manage to maintain their scores, since there is a clear degradation in their initial MAP values for BER values higher than 10^{-5}. CEDD is the most stable of the compilation, since the MAP value degrades only by 2.82 % compared to its initial value. For the rest of the global descriptors evaluated, Tamura appears to have the most degradation among the others, a total of 81.35 %, while at the same time it obtains the lowest initial score. CLD, being the smallest descriptor of this

DOI: 10.1002/dac

compilation, degrades by 16.36 % compared to its initial value (BER=0) for BER = 0.01. Figure 4 presents Precision @ 4 values for the experiments performed on the Nister database.

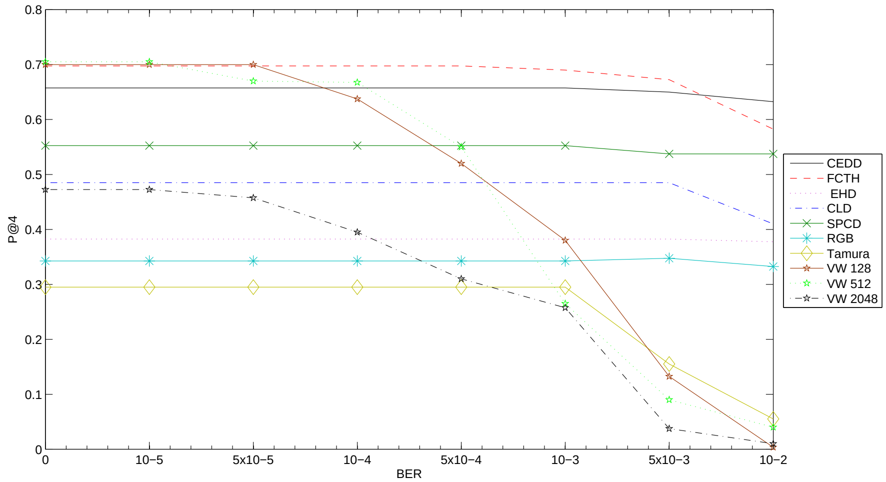

Figure 4. Nister Database: Precision@4 vs BER

By comparing Figures 3 and 4, we clearly observe that P@4 values are highly correlated to MAP values. Our conclusions and observations are common to those previously mentioned, since it is obvious that those descriptors maintaining high MAP values, thus being able to perform image retrieval, are capable of retrieving the desired images in the early 4 positions. Moreover, we must emphasize on the fact that the majority of the descriptors maintain their initial scores due to their small size in bytes. A small size packet, such as a descriptor, is more robust to errors since only a very small portion of the total amount of bits is affected by errors, especially for relatively low BER values. This assumption is confirmed by observing the performance of Visual Words, since due to their large size (> 1000 Bytes) compared to the rest of the descriptors (9 - 256 Bytes), their performance degrades rapidly even for low BER values.

Experimental results on the same scenario, performed on the Wang database, are presented in Figures 5 and 6. In Figure 5 we present Mean Average Precision, when the descriptor is transmitted through the network. The behavior of each descriptor is similar to the one observed in Figure 3. That is, almost all of the descriptors maintain their initial scores up to a BER value of 10^{-3}. Only VW-512 maintains its initial performance up to a relatively low BER value 10^{-4}, while at BER=10^{-3} it presents a degradation of 27.79 % compared to its initial MAP value. VW-512 has the worst performance compared to the others, with a total degradation of 74.3 %. As of the Compact Composite Descriptors, they maintain their initial MAP values for all the examined BER values, having a degradation of less than 5 %. The same observation can be made about the CLD descriptor, being the one with the smallest size among our evaluation. A difference compared to the Nister database is that Visual Words have lower initial MAP values which they maintain for larger BER values. As mentioned earlier, the Wang database consists of query images where the ground truth contains images with similar visual content to the query image. Due to the nature of the Wang database, global features behave better than visual words [25].

DOI: 10.1002/dac

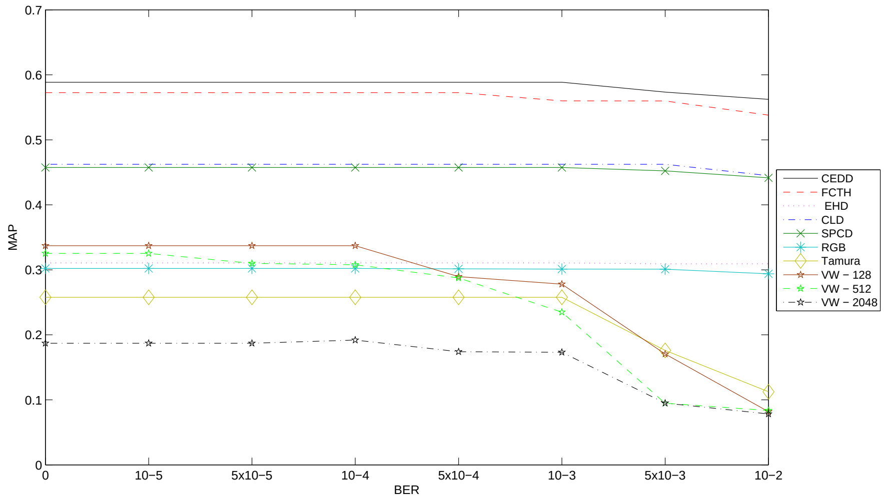

Figure 5. Wang Database:MAP vs BER

The results of Precision @ 10 for the Wang database are presented in Figure 6. As in the Nister database, also in this case there is a high degree of correlation between the results on Figures 5 and 6. What is really impressive in Figure 6 is the fact that VW-128 has a rapid degradation of its performance, reaching an almost zero value at BER=10^{-2}. It is obvious that we didn't expect to get a high value in this case, yet it is the only zero value observed among our compilation. Our conclusion is similar to the one for the Nister database, that is descriptors which maintain high MAP values, effectively performing image retrieval, are capable of retrieving the desired images in the early 10 positions. Also, the small size of the data files transmitted over the network results in less bits affected by errors, as mentioned for the Nister database results.

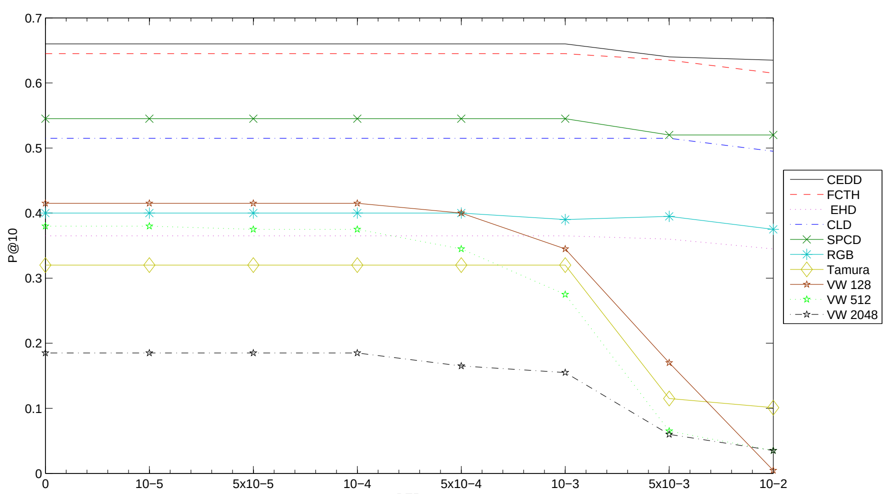

Figure 6. Wang Database: Precision@10 vs BER

DOI: 10.1002/dac

4.2. Transmitting the query image

In this setup the mobile user A transmits the query image to the Access Point through the wireless channel; therefore a larger amount of data is transmitted through the network. This increase in the amount of transmitted data has an effect on the number of bit errors introduced during the transmission due to the wireless channel's BER, since a large file (image) is much prone to errors compared to a smaller one (descriptor).

Figure 7. Image Corruption Under Different BER Levels

We use images compressed according to the JPEG 2000 standard. The reason why we did not use JPEG images to perform image retrieval is due to their error sensitivity. A JPEG image could be completely corrupted even with a very small number of corrupted (reversed) bits, making JPEG images inappropriate for our purpose. Contrary to JPEG, JPEG 2000 images are robust to bit errors, since the level of a, being appropriate for our scope. Our goal in this setup is to evaluate a descriptor's ability to perform image retrieval based on an image corrupted by errors during its transmission over a wireless channel. Figure 7 illustrates the visual corruption of an image under different BER levels. Please note that this image belongs to the Wang database.

Additionally, Figure 8 illustrates the first 3 retrieved images in case of corrupted queries. Query image is also part of the database. In all cases, images were retrieved using the CEDD descriptor. In the case of BER=0, one can see the first result is the image itself. The rest of the images belong to the same semantic group, i.e. ancient monuments. In case of BER=0.0005, only the first result is relevant to the query. The rest 2 retrieved images are from a completely different semantically set. Final, in case of BER=0.01, CEDD descriptor has failed to retrieve the query image from the database. All the three retrieved images are semantically-different to the query.

DOI: 10.1002/dac

Figure 8. Retrieval Results Under Different BER Levels

From Figure 9 (MAP for the experiments performed on the Nister database) it is obvious that image retrieval based on a corrupted image is much harder than using a corrupted descriptor. This is due to the fact that a packet of greater size is transmitted through the network, thus being more vulnerable to bit errors, affecting the overall performance of the retrieval process. The relationship of BER and optimum packet size, as long as the impact of BER and packet size have on network performance have already been discussed in [48]. All descriptors lose their initial MAP scores for BER values higher than 10^{-5}. In terms of MAP values, Tamura has the lowest initial MAP score, 0.3, and the lowest score, equal to 0.001 for BER = 10^{-2}. Its performance is degraded by almost 100 %, therefore it is completely inefficient and inappropriate to be used in such a scenario. Also in this scenario, CLD appears to maintain its shape. It has the third best MAP value at BER =10^{-4} and the higher among the compilation for BER values higher than 10^{-3}.

Among the Compact Composite Descriptors, FCTH attains the higher initial MAP score, for BER = 0, directly comparable to the scores of Visual Words. CEDD loses much of its retrieval capability for BER values higher than 10^{-4}, although it appears to have an initial MAP value of 0.6937. CEDD also attains the lowest MAP value between the others compact composite descriptors. SpCD on the other hand, despite its initial MAP value of 0.6128 which is lower than the rest compact composite descriptors, is the one having the highest MAP value compared to the other descriptors of the same category, for BER values higher than 5×10^{-4}. Visual Words have the highest gradient score among all other descriptors, i.e. their performance deteriorates more rapidly as BER increases.

Prepared using dacauth.cls
DOI: 10.1002/dac

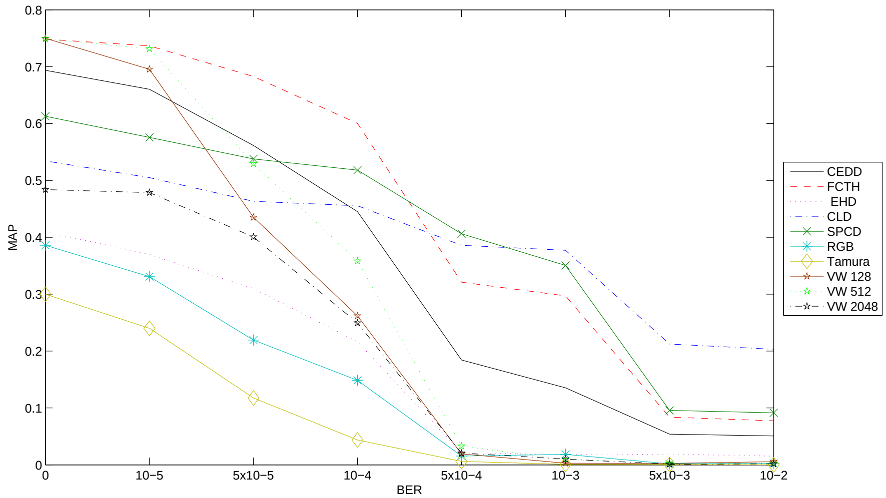

Figure 9. Nister Database: MAP vs BER

Their performance is better compared to other descriptors for low BER values, up to 5×10^{-5} however it deteriorates as BER values increase and gets inferior compared to compact composite descriptors and CLD. MAP values for Visual Words are very low, almost equal to zero, for BER values higher than 5×10^{-4}, thus being incapable of performing image retrieval.

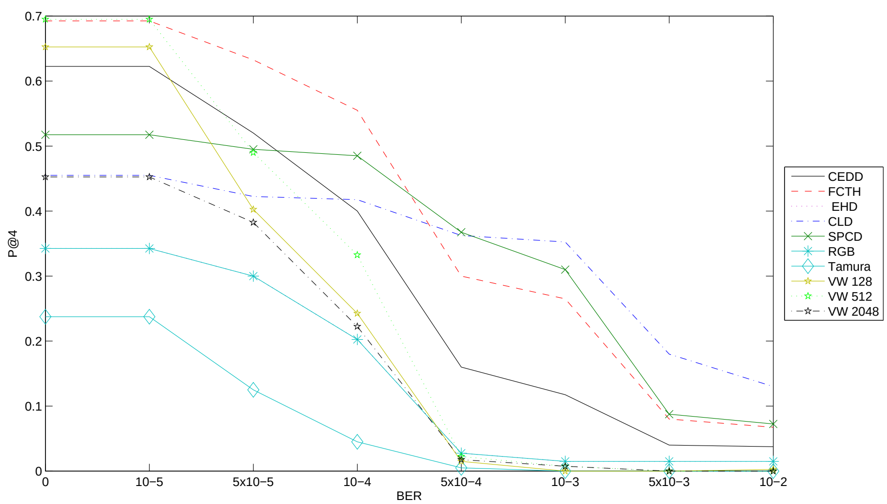

Figure 10. Nister Database: Precision@4 vs BER

By observing results of P@4 for different BER values on Figure 10, our conclusion is that the overall behavior is similar to the one observed on Figure 9, for MAP values. Also in this case CLD attains the highest P@4 scores for BER values higher than 5×10^{-4}, with SpCD being the second best for such BER values. For BER = 0, FCTH appears to have the greater P@4 score, directly comparable to the score of VW - 512. There still is the threshold of 5×10^{-4} after which VWs have very low scores. Finally, Tamura has the lowest initial and final P@4 values among the rest of the descriptors.

Corresponding results on the Wang database are presented on Figures 11 and 12. MAP results for the case where the image is transmitted through the network are presented in Fig.9. The performance of the descriptors under consideration for the Wang database, is similar to that observed in Figure 6 for the Nister database. In this case, descriptors appear to be relatively more robust compared to the Wang database scenario, since most of them, except from CEDD and Tamura, lose their initial MAP scores for BER values higher than 5×10^{-5}. The lowest initial value is observed for VW-2048, whereas CEDD which has the highest initial value appears to also have the greatest initial value degradation equal to 72.41 %. There may be a number of differences to the MAP score ranking of the descriptors compared to the corresponding results of the Nister database, however the overall systems behavior is similar in both cases.

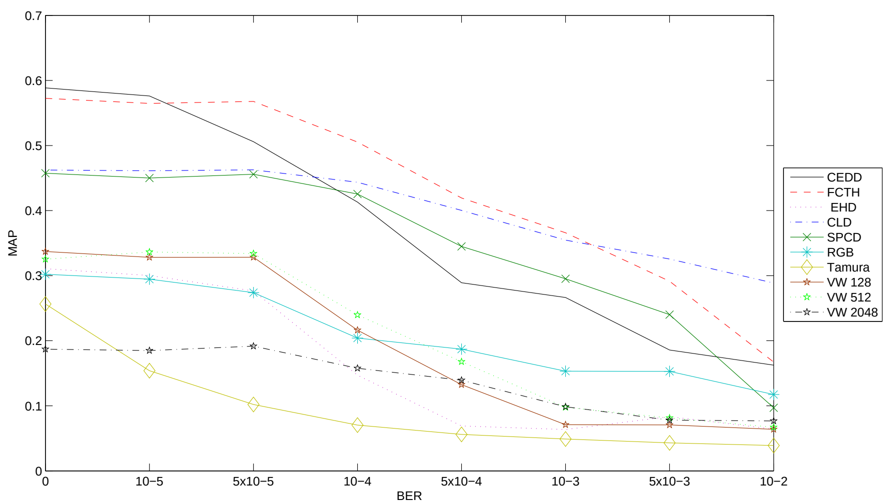

Figure 11. Wang Database: MAP vs BER

Concluding, our evaluation attempt shows that it is preferable in terms of effectiveness to locally extract a descriptor than send the image itself for the purposes of image retrieval. Such an operation, today, is easily performed, given that the computational power of a mobile device allows it. In any case, even though we do not suggest the image transmission scenario, based on the aforementioned results we conclude the following:

CLD has the best performance mainly because of its structure. In order to explain the behavior of CLD it is necessary to investigate its structure. CLD separates an image into 64 blocks and calculates the average color on each one of them. The noise corrupting an image, due to this segmentation, is spread among the image, thus not affecting severely the average color value of each block.

An image corrupted by noise is seriously affected in terms of its texture. Given that CEDD and FCTH are strongly related to texture information, we observe a rapid performance degradation. The same factor affects Tamura and EHD descriptors.

DOI: 10.1002/dac

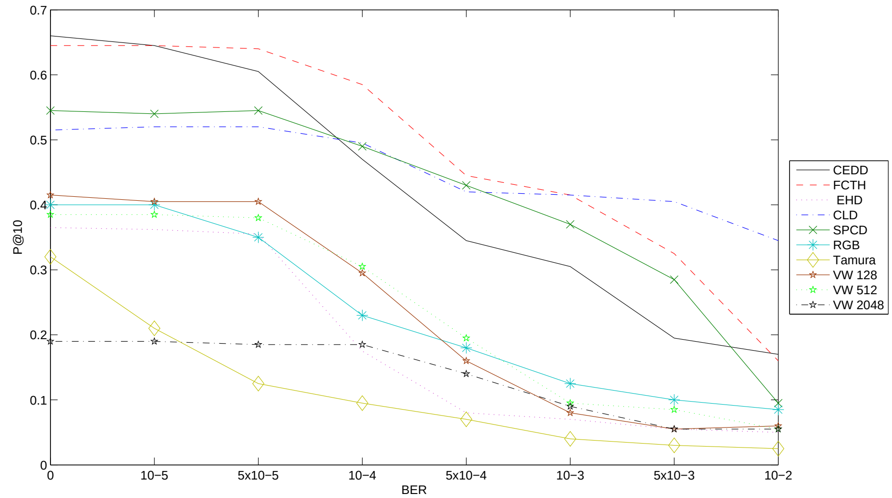

Figure 12. Wang Database: Precision@10 vs BER

The addition of noise to an image may create new points of interest without any ‘interest’. Therefore, since VWs use points of interest to describe the content of an image, the existence of false points affects the accuracy of the results.

4.3. *Descriptor’s Size and Performance Degradation*

Moreover, we want to examine the tradeoff between a descriptor’s size and the degradation of its performance. We implement this comparison by considering another metric named the Normalized MAP, N-MAP, which represents the MAP value per Byte for every descriptor. In order to calculate this metric we take into account the total number of Bytes, both correct and erroneous. The results are presented in Figure 13, for BER=0.01. We used all the results from the descriptor transmission setup for the Nister database, and the results for CLD descriptor from the image transmission setup also for the Nister database, since it has the best performance among the others.

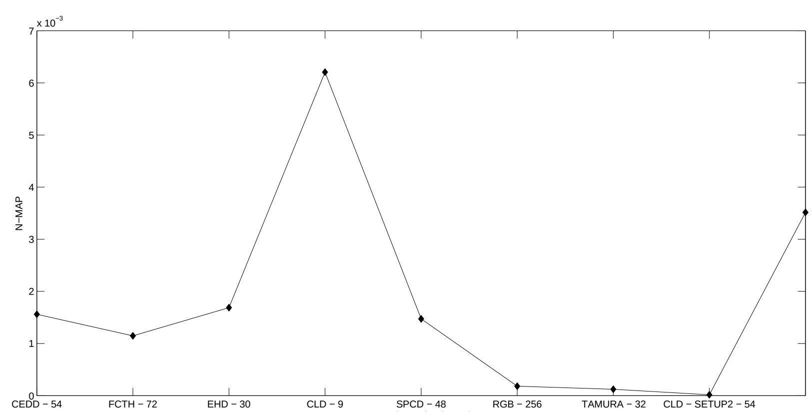

Figure 13. N-MAP value - Tradeoff Between Descriptor Size and Performance Degradation

DOI: 10.1002/dac

As shown in Figure 13, CLD descriptor has the highest N-MAP value, showing that despite its small size and moderate MAP value compared to all other descriptors, each of its Byte carries the greatest amount of valid information. But, in any case, we do not propose the use of CLD descriptor, since the N-MAP metric is calculated only to demonstrate the performance degradation in relation with a descriptor's size. In a real world scenario it is more important to achieve more accurate results even if it necessary to transmit higher amount of data. For this reason, we CEDD appears to be the most robust in terms of noise tolerance among the compilation

5. CONCLUSIONS

The main contribution in this paper is that not only did we examine image retrieval on a larger and much more realistic network, but also performed content based image retrieval process using compressed images, such as JPEG2000 on both Nister and Wang databases. We conduct an evaluation of the image retrieval procedure over a wireless noisy channel of an IEEE 802.11 network, consisting of an AP and a number of wireless nodes. Our goal is t oinvestigate a descriptor's resilience to noise, by conducting experiments on two different configurations for each database. In the first one, descriptors are extracted locally to the mobile device and then transmitted over a noisy wireless channel, thus arriving corrupted at the AP. The AP in turn forwards the descriptor to a remote server where content based image retrieval is conducted. In the second configuration, a query image compressed according to the JPEG 2000 standard is transmitted through the network.

During the first setup, the majority of the descriptors manage to maintain high retrieval scores for BER values up to 0.01, except from Visual Words and Tamura, whose performance degrades severely for BER values higher than 1×10^{-4} and 1×10^{-3} respectively. In the second configuration, our results show all descriptors have their MAP values degrading as BER values increase. In this case, CLD appears to have the best performance for high BER values, whereas Visual Words maintain average MAP scores for BER values greater than 5×10^{-4}.

Our main conclusion drawn is that it is preferable to send a descriptor than an image over the network for two main reasons: the image retrieval process is more accurate and the overall network load is significantly lower. Additionally we demonstrated that using compact descriptors such as compact composite descriptors and MPEG-7 descriptors is much more effective, since retrieval effectiveness is directly correlated with a descriptor's size and storage requirements. Overall, for the examined configuration CEDD appears to be the most robust in terms of noise tolerance among the compilation.

It is our intention in a future work to examine the portion and place of bits of a descriptor (or image) which mostly affect content based image retrieval process. Since IEEE 802.11 MAC is a contention-based wireless MAC, we plan to examine the impact contention has on the system's performance, by increasing the number of nodes and applications, on the image retrieval process. Moreover, we will investigate the effect of multiple data flows on the image retrieval process, in terms of delay, fairness etc. and suggest improvements, as done in [49]. Furthermore, we plan to extend the experiments and to evaluate the performance of other compact descriptors from the recent literature such as [50] and [51]. Additionally, future experiments will include the evaluation of image retrieval when multimodal data are employed e.g. image and text [52].

DOI: 10.1002/dac

ACKNOWLEDGMENT

This work is elaborated through an on-going THALES project (CYBERSENSORS - High Frequency Monitoring System for Integrated Water Resources Management of Rivers). The Project has been co-financed by the European Union (European Social Fund - ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: Thales. Investing in knowledge society through the European Social Fund.

REFERENCES

1. E. Chatzistavros, S. A. Chatzichristofis, G. Stamatellos, K. Zagoris, Comparative performance evaluation of image descriptors over ieee 802.11b noisy wireless networks, in: MUE, IEEE, 2011, pp. 128–133.
2. M.-H. Ho, C.-C. Chiu, S.-H. Liao, Comparison of different antenna arrays for the ber reduction in indoor wireless communication, International Journal of Communication Systems 26 (2) (2013) 161–176.
3. J. Postel, RFC 768: User Datagram Protocol. Internet Engineering Task Force, 1980.
4. B. Girod, V. Chandrasekhar, D. M. Chen, N. . Cheung, R. Grzeszczuk, Y. Reznik, G. Takacs, S. S. Tsai, R. Vedantham, Mobile visual search, IEEE Signal Processing Magazine 28 (4) (2011) 61–76.
5. I. Ahmad, M. Gabbouj, A generic content-based image retrieval framework for mobile devices, Multimedia Tools Appl. 55 (3) (2011) 423–442.
6. R. He, K. Liu, N. Xiong, Y. Zhu, Garment image retrieval on the web with ubiquitous camera-phone, in: APSCC, IEEE, 2008, pp. 1584–1589.
7. V. Chandrasekhar, S. S. Tsai, G. Takacs, D. M. Chen, N. . Cheung, R. Vedantham, Y. Reznik, R. Grzeszczuk, B. Girod, Low latency image retrieval with progressive transmission of chog descriptors, in: MCMC'10 - Proceedings of the 2010 ACM Workshop on Mobile Cloud Media Computing, Co-located with ACM Multimedia 2010, 2010, pp. 41–46.
8. V. Chandrasekhar, D. M. Chen, A. Lin, G. Takacs, S. S. Tsai, N.-M. Cheung, Y. A. Reznik, R. Grzeszczuk, B. Girod, Comparison of local feature descriptors for mobile visual search, in: ICIP, IEEE, 2010, pp. 3885–3888.
9. A. Nakagawa, A. Kutics, K. H. Phyu, Mosir: Mobile-based segment and image retrieval, in: 7th International Conference on Wireless Communications, Networking and Mobile Computing, WiCOM 2011, 2011.
10. IEEE, IEEE standard for information technology–telecommunications and information exchange between systems local and metropolitan area networks–specific requirements part 11: Wireless lan medium access control (mac) and physical layer (phy) specifications, IEEE Std 802.11-2012 (Revision of IEEE Std 802.11-2007) (2012) 1 – 2793doi:10.1109/IEEESTD.2012.6178212.
11. T. Issariyakul, E. Hossain, Introduction to Network Simulator NS-2, Springer, 2009.
12. P. Mahasukhon, M. Hempel, H. Sharif, T. Zhou, S. Ci, H.-H. Chen, Ber analysis of 802.11b networks under mobility, in: ICC, IEEE, 2007, pp. 4722–4727.
13. R. Misra, C. R. Mandal, Performance comparison of aodv/dsr on-demand routing protocols for ad hoc networks in constrained situation, in: IEEE International Conference on Personal Wireless Communications, 2005, pp. 86–89.
14. D. Comer, Internetworking with TCP/IP, Prentice Hall, 2001.
15. R. Dunaytsev, D. Moltchanov, Y. Koucheryavy, J. Harju, Modeling tcp sack performance over wireless channels with completely reliable arq/fec, Int. J. Communication Systems 24 (12) (2011) 1533–1564.
16. R. Dunaytsev, D. Moltchanov, Y. Koucheryavy, J. Harju, Modeling tcp sack performance over wireless channels with completely reliable arq/fec, International Journal of Communication Systems 24 (12) (2011) 1533–1564.
17. M. Patzold, J. Wiley, Mobile fading channels, John Wiley, 2002.
18. L. Greenstein, D. Michelson, V. Erceg, Moment-method estimation of the ricean k-factor, Communications Letters, IEEE 3 (6) (1999) 175 –176. doi:10.1109/4234.769521.
19. I. Sarris, A. Nix, Ricean k-factor measurements in a home and an office environment in the 60 ghz band, in: Mobile and Wireless Communications Summit, 2007. 16th IST, 2007, pp. 1 –5. doi:10.1109/ISTMWC.2007.4299295.
20. R. Datta, D. Joshi, J. Li, J. Wang, Image retrieval: Ideas, influences, and trends of the new age, ACM Computing Surveys 40(2) (2008) 1–60.
21. M. Lux, S. A. Chatzichristofis, Lire: lucene image retrieval: an extensible java cbir library, in: ACM Multimedia, 2008, pp. 1085–1088.

22. D. P. Papadopoulos, V. S. Kalogeiton, S. A. Chatzichristofis, N. Papamarkos, Automatic summarization and annotation of videos with lack of metadata information, Expert Systems with Applications 40 (14) (2013) 5765 – 5778. doi:http://dx.doi.org/10.1016/j.eswa.2013.02.016.
23. R. Ren, J. P. Collomosse, J. M. Jose, A bovw based query generative model, in: K.-T. Lee, W.-H. Tsai, H.-Y. M. Liao, T. Chen, J.-W. Hsieh, C.-C. Tseng (Eds.), MMM (1), Vol. 6523 of Lecture Notes in Computer Science, Springer, 2011, pp. 118–128.
24. S. A. Chatzichristofis, K. Zagoris, A. Arampatzis, Bag-of-visual-words vs global image descriptors on two-stage multimodal retrieval, in: W.-Y. Ma, J.-Y. Nie, R. A. Baeza-Yates, T.-S. Chua, W. B. Croft (Eds.), SIGIR, ACM, 2011, pp. 1251–1252.
25. S. Chatzichristofis, C. Iakovidou, Y. Boutalis, O. Marques, Co.vi.wo.: Color visual words based on non-predefined size codebooks, Cybernetics, IEEE Transactions on 43 (1) (2013) 192–205. doi:10.1109/TSMCB.2012.2203300.
26. N. Pinto, D. D. Cox, J. J. DiCarlo, Why is real-world visual object recognition hard?, PLoS Computational Biology 4 (1).
27. Y.-L. Boureau, F. Bach, Y. LeCun, J. Ponce, Learning mid-level features for recognition, in: CVPR, 2010, pp. 2559–2566.
28. S. Lazebnik, C. Schmid, J. Ponce, Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories, in: CVPR (2), 2006, pp. 2169–2178.
29. J. Yang, Y.-G. Jiang, A. G. Hauptmann, C.-W. Ngo, Evaluating bag-of-visual-words representations in scene classification, in: Proceedings of the international workshop on Workshop on multimedia information retrieval, MIR '07, ACM, New York, NY, USA, 2007, pp. 197–206. doi:10.1145/1290082.1290111.
URL `http://doi.acm.org/10.1145/1290082.1290111`
30. A. Bolovinou, I. Pratikakis, S. Perantonis, Bag of spatio-visual words for context inference in scene classification, Pattern Recognition.
31. J. C. Niebles, H. Wang, L. Fei-Fei, Unsupervised learning of human action categories using spatial-temporal words, International Journal of Computer Vision 79 (3) (2008) 299–318.
32. Q.-F. Zheng, W.-Q. Wang, W. Gao, Effective and efficient object-based image retrieval using visual phrases, in: ACM Multimedia, 2006, pp. 77–80.
33. D. Kragic, H. I. Christensen, Robust visual servoing, The international journal of robotics research 22 (10-11) (2003) 923–939.
34. J. J. Steil, F. Röthling, R. Haschke, H. Ritter, Situated robot learning for multi-modal instruction and imitation of grasping, Robotics and Autonomous Systems 47 (2) (2004) 129–141.
35. S. A. Chatzichristofis, Y S Boutalis and Mathias Lux, Img(rummager): An interactive content based image retrieval system, in: 2nd International Workshop on Similarity Search and Applications (SISAP), 2009, pp. 151–153.
36. S. A. Chatzichristofis, K. Zagoris, Y. S. Boutalis, N. Papamarkos, Accurate image retrieval based on compact composite descriptors and relevance feedback information, IJPRAI 24 (2) (2010) 207–244.
37. S. Chatzichristofis, Y. Boutalis, Cedd: Color and edge directivity descriptor: A compact descriptor for image indexing and retrieval, Vol. 5008, Springer, 2008, p. 312.
38. S. A. Chatzichristofis, Y. S. Boutalis, Fcth: Fuzzy color and texture histogram - a low level feature for accurate image retrieval, in: WIAMIS, 2008, pp. 191–196.
39. B. Manjunath, J. Ohm, V. Vasudevan, A. Yamada, et al., Color and texture descriptors, IEEE Transactions on circuits and systems for video technology 11 (6) (2001) 703–715.
40. S. A. Chatzichristofis, Y. S. Boutalis, M. Lux, Spcd - spatial color distribution descriptor - a fuzzy rule based compact composite descriptor appropriate for hand drawn color sketches retrieval, in: J. Filipe, A. L. N. Fred, B. Sharp (Eds.), ICAART (1), INSTICC Press, 2010, pp. 58–63.
41. H. Tamura, S. Mori, T. Yamawaki, Textural features corresponding to visual perception, IEEE Transactions on Systems, Man and Cybernetics 8 (6) (1978) 460–473.
42. T. Deselaers, D. Keysers, H. Ney, Features for image retrieval: an experimental comparison, Information Retrieval 11 (2) (2008) 77–107.
43. O. G. Cula, K. J. Dana, Compact representation of bidirectional texture functions, in: CVPR (1), IEEE Computer Society, 2001, pp. 1041–1047.
44. H. Bay, T. Tuytelaars, L. J. V. Gool, Surf: Speeded up robust features, in: ECCV (1), 2006, pp. 404–417.
45. D. Nistér, H. Stewénius, Scalable recognition with a vocabulary tree, in: CVPR (2), IEEE Computer Society, 2006, pp. 2161–2168.
46. J. Z. Wang, J. Li, G. Wiederhold, Simplicity: Semantics-sensitive integrated matching for picture libraries, IEEE Trans. Pattern Anal. Mach. Intell. 23 (9) (2001) 947–963.
47. K.-M. Wong, K.-W. Cheung, L.-M. Po, Mirror: an interactive content based image retrieval system, in: Circuits and Systems, 2005. ISCAS 2005. IEEE International Symposium on, 2005, pp. 1541 – 1544 Vol. 2.

doi:10.1109/ISCAS.2005.1464894.
48. E. Modiano, An adaptive algorithm for optimizing the packet size used in wireless arq protocols, Wireless Networks 5 (4) (1999) 279–286.
49. Y. Xiao, F. H. Li, M. Li, J. Zhang, B. Li, Dynamic bandwidth partition schemes for integrated voice, video, and data traffic in the ieee 802.11e distributed wireless lans, Int. J. Communication Systems 23 (3) (2010) 391–412.
50. S.-M. Park, K.-J. Kim, Color recognition with compact color features, Int. J. Communication Systems 25 (6) (2012) 749–762.
51. K. Zagoris, E. Kavallieratou, N. Papamarkos, Image retrieval systems based on compact shape descriptor and relevance feedback information, J. Visual Communication and Image Representation 22 (5) (2011) 378–390.
52. A. Arampatzis, K. Zagoris, S. A. Chatzichristofis, Dynamic two-stage image retrieval from large multimedia databases, Inf. Process. Manage. 49 (1) (2013) 274–285.

DOI: 10.1002/dac