A NEW APPROACH FOR RASTER IMAGES ANNOTATION

Giruta Kazakeviciute - Januskeviciene1, Eugenijus Januskevicius2

1Vilnius Gediminas Technical University, Department of Graphical Systems, 
Sauletekio al. 11, Vilnius, Lithuania, giruta@vgtu.lt
2Vilnius Gediminas Technical University, Faculty of Architecture, 
Pylimo g. 26/1, Vilnius, Lithuania, ejs@ar.vgtu.lt

Abstract. Growth of media data creates huge problems when searching and classifying the depicted objects. Currently search is based either on surrounding context, which is irrelevant in most cases, provided image meta data sets or low or mid-level regional peculiarities of digital images. The article is focused on a specific case of image meta data - annotations: textual elements, describing or explaining objects, depicted in the image. The paper presents novel approach of data storage in the field of digital imaging, embedding annotation data in format independent way directly into digital images. The annotation scheme presented is targeted to JPEG-2000 as upcoming image storage and transmission standard, but can be easily incorporated into other image storage formats, both lossless and lossy. The scheme allows wide taxonomy of images to be annotated: digital photos, aerial photos, scanned material, medical data sets, etc. The scheme keeps visual and structural fidelity of digital images, exploring properties of human visual system and wavelet transformation. The primary aims of the presented scheme is sharing and transfer of knowledge, like preparing material by educators, but can be extended to broader fields of application, like video, data mining or replacement for image tags and descriptions, thus creating true searchable media environment.

Keywords: annotation, data embedding, digital image, fidelity.

1 Introduction

Nowadays a large and growing amount of information is stored in various multimedia formats – as images, video, audio. A common and successful approach to organize and manage huge quantities of information is to enrich the media with meta data, tags, labels and annotations. Textual information attached is not only a way to organize information, but it enables the creation of a knowledge base or knowledge base systems for different applications, can be used for knowledge discovery and information mining from task-based imagery. It is said: “We are drowning in information but starved for knowledge ” (John Naisbitt).

Image annotation is multidisciplinary area and related with computer vision, data processing, data mining (also called knowledge discovery and data mining – KDD), artificial intelligence, machine learning and more others. It is widely acknowledged that image annotation is an open and very difficult problem in information technology science. Image annotations methods depends on images application, on image type and format, on image storage repository.

There are numerous users for which precise image annotation is very important: researchers, medics, students storing images on personal computers hard drives, ordinary users inside an organization, etc. These users require not only an accurate description of the image, like author, date, time etc., contained in meta data, but also domain specific knowledge. For this purpose the annotations should be performed in a standardized and constrained manner. Typically, in [2], [5], Error: Reference source not found, [11] content has been represented using a strict ontology to ensure that the information is consistent. This is the most common type of annotation in the semantic web. Ontology for annotations can be quite complex and time-consuming. Most of the current annotation tools provide the ontology, which is displayed in the form of a large tree. Annotation is done by scrolling over a tree and selecting an element from this tree.

In other cases, the strict constraints are less relevant, for example users that simply want to annotate their own pictures for sharing with friends [12], [13]. For personal use folksonomies – free annotation are more popular. It is a way to attach emotions and memories. It is convenient when annotation of images can happen every time an image is accessed. Naturally, not all annotations must be widely available. Some annotations can be personal, others may stay within specific boundaries (an example, in a department), and others can be made available. The most frequent application, analyzed in science literature is image retrieval [2], [11], [12], sometimes management of personal digital photo collections Error: Reference source not found and medical imagery [4], [5], [9], Error: Reference source not found, rarely education, data mining, transfer of knowledge [4] and similarly, and these tasks are analyzed on the semantic WEB environment.

The primary focus of this work – on the novel approach of data storage in the field of digital imaging, embedding annotation data in format independent way directly into digital images with the purpose sharing and transfer of knowledge not only on line but off line too.

Section “State-of-the Art” presents the existing image annotation methods, thus underlying the novelty
of our approach. After description of general requirements for image annotation, the quality evaluation metric for annotated images is proposed. In section “A new annotation strategy for task-based imagery” is reviewed the new image annotation scheme. The article closes with experimental results, conclusions and future work.

2 State-of-the-Art

2.1 Methods for images annotation

The widely used paradigm for information description in digital media is the use of supplementary data – meta data. It is usually organized in pairs of attribute and value, and to some extent is self-describing. Text-based image annotation continues to be an important fundamental problem in the computer vision, information mining, retrieval communities. With rapidly increasing collections of image data on and off the Web, image context understanding, search and retrieval is becoming actual an employment of proper annotating methods for task-based imagery. The most popular image annotating methods is presented below.

2.1.1 Dissociated annotations

In the majority of works [4], [5], [9], author or user can add information to an image using Extensible Markup Language (XML), Resource Description Framework (RDF), Web Ontology Language (OWL).

XML-based annotations play an important role in describing on line images information on the Internet. XML-based annotation information is stored independently from the image data, and combined with the image data at display time. An “annotated image” is the combination of an “annotation file” and its associated image. A single image may be referenced by many annotation files. The description in RDF form can include any meta data, defined in XML scheme.

Ontology is used as the new mean of creating and using the meta data in annotation, search and retrieval. It basically contains concepts and their relationships and rules. Adding a hierarchical structure to a collection of keywords produces a taxonomy. An ontology can solve the problem when some keywords are unclear. Ontologies are important for the Semantic Web, and a number of languages, such as OWL and RDF, exist for their formalization. Formats like DIG35 [32] or MPEG7 [33] are milestones to create and handle images descriptions in a flexible way. They are based on XML-description scheme [34] and believed to be the best paradigms to handle structured data. Here information is described and organized by using extendable, hierarchical and freely-definable structures.

Despite the good properties of XML-based approaches, annotations of such type can be lost if the image file format or image name is changed or user is off line. When the image is transmitted XML-based annotation information create additional payload.

2.1.2 Annotations embedding in to the image format

The application of meta data is closely related with the media storage format, as usually the storage format defines the options for meta data storage and retrieval. Basic implementations of this paradigm can be found in applications where the available information for the description is supposed to be fixed and non-expandable. Typical examples are the EXIF information, used in digital cameras, or Dublin Core Meta data standard [15]. This standard is used to add meta data to a wide variety of resources in a simple manner. Similar to the Dublin Core categories, but focusing only on images, are the Visual Resources Association (VRA) Core Categories [25].

Embedding information into an image file format can be done in a variety of ways. The simplest is the use of features provided by the file format, like JPEG (EXIF) [6] or PNG image file formats. Fixed fields are reserved for attribute values and might be filled with the current description, whereby quantity and kind of attributes depend directly on the file format used. Another and more flexible approach is to embed arbitrary meta data is used in modern formats like JPEG2000 [7] or in parts in the TIFF or TGA formats [8]. Here the concept of containers or boxes to store the information is used. The number of boxes is not limited and they can be filled with arbitrary content, e.g. user-defined attribute pairs.

However, all those approaches simply attach additional data to the image, creating additional payload when the image is transmitted. As the attached information is not integral part of the image data, there is no uniform support for certain attributes in every image file format, and most of the description is lost if the file format is changed.

2.2 Information embedding in to the image content

The idea is related with information hiding techniques, such as watermarking and steganography. Traditionally, steganography have been designed for security-related applications and watermarking, like
copyright protection and data authentication. Recently, researchers have attempted to broaden information hiding application scope to other oriented applications [35], [36], [1], [37].

2.2.1 Visible annotations

Visible annotations as visible watermarks is the standard part of the image data but they clutter the image data and decrease visibility of the depicted objects on the image. These type of annotations is the most suitable for large images, like panoramic, satellite and similar [14]. There has been a plurality of literature focusing on label placement on maps beginning as early as 1962 with [38] and continuing today. Many systems introduce idea hiding objects when they are below a certain minimal magnification threshold.

2.2.2 Invisible annotations

The advantages of invisible annotation approach are that the embedded data becomes an inseparable part of the media and takes almost no additional storage space.

Many algorithms applied in steganography might also be suitable for embedding of meta data, since the general focus to emphasize capacity is one of main requirements for media description. Nevertheless, these approaches might also be adapted to further increase capacity by omitting the requirement of undetectability. Many of these approaches are designed for the spatial domain as it provides more capacity and controllability for embedding and extraction of the additional information. Here, the most common approach to embed data is to substitute certain bits of the available pixel representation by bits of the additional data [39]. This technique is known as LSB-approach in steganography.

Since it capacity is inversely proportional to robustness [40], these approaches are highly vulnerable if the image is manipulated after embedding. Nowadays, image data is mostly used in compressed form to reduce storage space and transmission costs, which often includes a lossy transformation of the image content. Due to their special alignment, watermarking techniques are more robust against such manipulations than steganographic approaches [40]. Most of them have been designed for a certain transformation domain and embed the additional data (watermark) directly into the transformed image representation to be resistant against modifications imposed by the transformation itself. Dependent on the used transformation domain, e.g. Discrete Cosines Transform (DCT) [41] or Discrete Wavelet Transform (DWT) [42], or specific file formats using transformation domains, e.g. JPEG Error: Reference source not found or JPEG2000 Error: Reference source not found, many different techniques have been proposed. Nevertheless, a general problem with all of these approaches is, the capacity is in general very low and the whole image is used for embedding to reduce the risk the embedded data can be simply removed by cropping parts of the image. To create a reasonable technique for media embedding relevant properties and approaches from steganography and watermarking approaches must be combined and enhanced.

A combined technique derived from steganography, is the LSB-approach applied to DWT coefficients. Bit plane modifications usually have the highest possible informational capacity but they are not robust to even the slightest modifications. As media description goal is to maximize capacity, we can adopt the LSB approach, keeping in mind its disadvantages. Given the general analyzes of current annotation methods, it is clear there is much room for improvement in the state-of-the-art.

2.3 General requirements for image annotation

As the focus is on embedding data directly into a host image, the requirements of approach are highly related with steganography and watermarking, but has some differences [1].

The goal of an image annotation – provide additional cognitive information to the information, already presented in the image. Typical use scenarios include:

- providing descriptive information on separate objects of the image;
- revealing image peculiarities or the image or peculiarities of the objects depicted;
- providing non-obstructive way of descriptive data presentation;

Typical usage fields include:

- professional users, like instructors, using annotations to present the material; medical practitioners using annotations to denominate peculiarities of medical images; researchers and scientists, using annotations to spread the revealed knowledge;
- non-professional users, like home uses, using annotations to reveal objects depicted.

The annotated image is as useful as the image data is useful, so image data, but not the embedded information must be protected. This means that an annotation, its length or quality can be sacrificed over the quality of the digital image.

As the format of image annotation is open, there is potential risk of annotation be removed or altered.
We do not assume annotation will ever be used to provide vital data, so any alteration or removal of annotation is allowed. To provide necessary level of security, data-level security methods, like digital signatures, may be used.

Since there is no limitation regarding the image content to be described, shape and number of image regions, and the length of each description are a prior only limited by the properties of the host image. It is obvious that the sum of all annotations cannot exceed its provided capacity. Nevertheless, this is a strong demand assuming an optimal and redundancy-free usage of the provided capacity, which is rather hard to fulfill for arbitrary application fields.

Based on this assumptions, we are able to derive basic requirements for embedding region-based image descriptions into raster images:

1. The scheme is targeted to off-line users, or users, having limited data connection capabilities. The limitation is imposed to be able to use annotated data in situations when connection an external database or network node is limited for security reasons, connection costs or lack of infrastructure.
2. All embedded data belonging to regions shown on screen must be retrievable without any additional transmission of any data.
3. The extraction should not depend on current image quality.
4. The need to transmit annotations separately must be avoided.
5. Annotations may contain any binary data.
6. Regions to be described may have any shape.
7. Number and length of embedded data is limited by the provided capacity and quality of the host image.

2.4 Quality evaluation of annotated images

A digital imaging system involves many components: acquisition and storage, image processing and analysis, compression, transmission, printing and display. Each of these components can influence the final digital image quality. Accordingly, with the rapid development of imaging systems, have been developed the metrics of images quality.

We are interested mostly in image degradation during image compression and processing (transformed image data), because it is in our scope of this work.

Because of the differences in the characteristics of image application and contents, image formats and users, existing objective natural image quality metrics does not provide satisfactory results. Standard quantitative image quality metrics, such as peak signal to noise ratio (PSNR) and mean squared error (MSE), are not directly related to human perception. For the image quality evaluation we need automated objective quality assessment methods that are guided by the human vision model in order to accurately reflect human perception.

In the most image processing environments human eyes are the ultimate receivers of images. For many years the subjective quality measurement Mean Opinion Score (MOS) has been used. However, subjective evaluation is usually too inconvenient, time-consuming and expensive in practice.

Many suggestions have been made to develop new objective image quality measurement, incorporating perceptual quality measures by considering human visual system (HVS) characteristics [16], [17], [18]. But in [16], [23] has been noticed that none of the complicated objective image quality metrics has shown any clear advantage over simple mathematical measures such as PSNR under strict testing conditions using different image distortions. Later, in [19], a consortium of experts, the video quality experts group (VQEG) [20] reported that none of the new proposed methods, tested under different configurations, were comparable to the “null mode,” a hypothetical model that predicts quality exactly. It means QA methods needed to improve further.

The perceptibility of image details depends on the sampling density of the image signal, the distance from the image plane to the observer, and the perceptual capability of the observer’s visual system [22]. The subjective evaluation of the images varies when these factors vary. A single-scale method (SSIM) is appropriate only for specific settings. Using multi-scale method (M-SSIM) is possible to incorporate image details at different resolutions.

The authors of [22] have made the quality assessment (QA) algorithms freely available to the research community [24]. This has allow other researchers to report their comparative results on widely accepted QA algorithms. In [30], [31] have presented a novel information theoretic criterion – visual information fidelity (VIF) criterion for natural scenes. VIF was developed for image and video quality measurement based on natural scene statistics (NSS). Images and videos of the three-dimensional (3-D) visual environment captured using high-quality capture devices operating in the visual spectrum are classified as natural scenes [30]. Images and videos captured from non-visual stimulus such as radar and sonar, X-rays, ultrasounds, computer generated graphics, cartoons and animations, paintings and drawings, random noise, etc. differ from natural scenes.

By combining information content weighting with multi-scale SSIM, in [26] has been defined an information content weighted SSIM measure (IW-SSIM). Authors of this research has looked at the image QA pooling problem from an information theoretic point of view. In computational vision science is widely
hypothesized that the human visual system (HVS) is an optimal information extractor [27]. To achieve this optimality, the image components that contain more information content would attract more visual attention [28]. The local information content can be quantified in units of bit. Such statistical image model is available and then the local information content measure can be employed for image QA weighting.

As we see from the recent extensive tests results with six publicly available independent image databases [29], the proposed IW-SSIM algorithm achieves the best overall performance.

3 A new annotation strategy for task-based imagery

The annotation scheme we are proposing is targeted towards communication and transmission of knowledge. We assume the possibility to annotate a depicted object is valued more than possibility to secure the annotated information. Because of this, security issues, like removing annotation or altering its contents are intentionally left aside.

The scheme (Fig. 1) proposed uses meta data to describe the spatial position of the Region of Annotation (ROA) in the image and textual or binary data to describe the contents of ROA (see Fig. 3).

The ROA can be of any geometrical shape. The only limitation set is ROA must fully be contained in the image or image part. If the same annotation is to be placed in different image parts, like tiles in JPEG 2000, every part must contain its own ROA. The coordinates and size of the shape are defined in spatial coordinate system of the initial image or image part. 3 basic shaped are defined now:

- rectangular, described by upper left corner, width and height;
- circular, described by center point and radius;
- free-shaped, described by a set of points, defining its boundary.

The data for annotation can be either textual, using Unicode as symbol set and allowing to embed any glyph, or binary, allowing any binary data, like image or sound, to be embedded into the carrier.

The scheme utilizes Laplacian pyramid decomposition, allowing compact annotation storage and tight spatial relation with the ROA (see Fig. 4). The annotation can be stored in the same spatial region as ROA or can be detached and stored elsewhere in the same image. There are two aims of annotation detaching:

- preservation of quality in ROA, needed in some image taxonomies, like medical images;
- increasing capacity of annotation, when it does not physically fit within ROA.

The proposed annotation scheme is targeted to high information density, so standard data description
schemes, like SGML or XML cannot be used. Instead of this, a hash, of “key-value” approach is used (see Fig. 6). This allows to store information in more compact form, because the total overhead of XML format for small data set can be as high as 3000% (see Fig. 2) but it tends to decreases while increasing the amount of useful information.

The annotation, containing both spatial information and annotating data, is packed into structure, suitable for embedding. The structure contains information of the annotation, graphical finder patterns and error-correcting codes. The graphical shape of the embedded structure reassembles Datamatrix - a family of 2D bar codes (see Fig 5).

The initial selection of embedding method is based on the lowest possible impact on information being annotated, low computational requirements and possibility to change the impact on the quality in the most flexible way. The embedding method chosen is a combination of LSB modifications and QIM and works in the lower bit planes of quantization indexes. Figure 7 shows bitplanes of HH subband before a) and b) and after embedding c) and d). Due to high entropy of the lower bit-planes, the embedded information is almost indistiguishable from the coefficient data.

The proposed scheme is targeted to JPEG 2000 image storage format and perfectly fits into standard JPEG 2000 image coding and decoding pipeline. Data and image preparation methods (see Fig. 1) for annotation are handled in cost-efficient way and can be re-used in any image storage format with negligible computational overload.

4 Experimental results

The proposed image annotation scheme was tested in laboratory conditions. To test annotation scheme, several publicly available image databases were used. Test images selected are from different categories, at least 3 images were used from each category. Testing goal was to measure visual impact on the image, caused by annotations. Only gray-scale images were tested, mainly because IW-SSIM index is defined for gray-scale images only.

Testing consists of 2 stages. The first stage consists of a single ROA, having size of 40-by-40 pixels. The payload for annotation is the first sentence from “Lorem ipsum...” containing 124 bytes. The second stage contains 5 annotations, each containing the same first sentence from “Lorem ipsum...”. The third stage contains 10 annotations, containing first 50 symbols from “Lorem ipsum...”. The results, presented in Figures 8 and 9 depicts average values of these tests.
IW-SIM, PSNR and MS-SIM values were measured. Additionally JPEG 2000 compression was introduced, varying compression ratios from 1:1 to 200:1. Maximal compression ratio taken was 200:1 for topographical images and 50:1 for medical images. Visual impact of embedded annotations was measured between the original uncompressed image and the compressed image containing annotations (set 1) as well as between compressed image and the compressed image containing annotations (set 2). The figures below presents average PSNR, IW-SSIM and MS-SSIM quality measurement for set 1 (left side) and 2 (right side) values for every type if image taxonomy tested. PSNR value is presented for reference only. The decision on image suitability for particular use should be made relying on either IW-SSIM, or MS-SSIM values. In this project we assume IW-SSIM value of 0.85 is the lowest suitable for natural and areal images, while 0.90 - for medical images in most modalities.

As presented by measurement series 2 in Figure 8, influence of moderate information embedding is almost unnoticeable for the user and image quality. The increase of image quality from compression ratio 1:1 to 1:2 or 1:4 should be explained by different DWT filters used. Average retrieval rate of embedded information in this stage is 70%. This can be improved by using more sophisticated methods of information embedding, as mentioned in “Future work” section.

5 Conclusions and future work

The proposed approach is suitable for information embedding in JPEG-2000 imaginary in small and medium quantities. The embedding process can be integrated in the process of JPEG-2000 encoding and take almost no additional computational time. Visual impact on the compressed image is neglectable when compared with the compression artifacts.

The analysis of retrieval process reveals the information is lost due to peculiarities of JPEG-2000 encoding process, when targeted bitrate is nearly reached. To improve the quality of information retrieval, a lifting scheme, used in ROI encoding could be used.

Color images can be used for information embedding, having suitable quality measurement methods. The targeted component is “Y”, or luminosity component. The choice is based on analysis of test results, as the quality of image does not suffer noticeably while embedding the information. Moreover, the Y component exposes just a little more entropy, as it is compressed less in comparison to U and V components.

References


* Some modalities does not allow DWT-based compression, while permitting DCT-based compression.


