

Embedding Semantic Information into the Content of Natural Scenes Images

G. Kazakeviciute-Januskeviciene¹, E. Januskevicius²

¹*Department of Graphical systems, Vilnius Gediminas Technical University, Saulėtekio av.11, Vilnius, Lithuania, phone: +370 5 2744848*

²*Department of Building Structures, Vilnius Gediminas Technical University, Pylimo St. 26/1, Vilnius, Lithuania; phone: +370 5 2745205
giruta@vgtu.lt*

Abstract—The article presents methodology for the annotation and semantic description of the natural images and their regions. The tools are based on information embedding directly into the JPEG 2000 image content and inherits standard properties of JPEG 2000. The influence of embedded information is evaluated using HVS-based quality metrics. The annotation methodology is oriented, but not limited, to off-line users.

Index Terms—Information embedding, image quality, JPEG 2000, annotation

I. INTRODUCTION

There is a number of computer based systems that facilitate to manage, browse and retrieve photo images [1]–[3], however, not so many for personal recollection of objects in traveling places, persons, events [4], [5] and only few systems for knowledge discovery, acquisition and learning [6]. One from them based on the image content [1], other – on context and remaining combine all available information (low level image features, time, geolocation, tags, annotations and other) [2], [3], [5].

Scientists, journalists, hobbyists and more others describe photography with a memorable phrases or explanatory notes, because not everything and not for everybody is clear in the content of digital photos. There are objects and their properties that are difficult to presume or guess. Naturally, we leave out what is visually obvious.

The valuable textual digital information of photo images, manually produced by the authors of the images or other interested persons, including information embedded by cameras, can be linked by one term – meta data. In current scope of the work only one type of meta data is called annotations – descriptive and/or explanatory information about the content of the photo.

The tools are created for embedding explicit semantic information about the real-word objects and their parts direct into the content of the photo image. There is no need for the special software to read descriptions, no troubles about file formats or names, no additional files and payloads. The suggested annotation methodology is oriented for offline

computer users, as well as are handy for online users. In some cases there are limited data connection capabilities, when connection to an external database or network node is limited for security reasons, connection costs or lack of infrastructure and web implementations does not work. Using the tools, anyone is able to modify or add annotations into the content of image in the most convenient way.

II. DESCRIPTIVE INFORMATION OF DIGITAL PHOTO IMAGES

In the age of digital photos, authors and users need information that can be stored with the file and would be portable.

The real-world objects semantic understanding in digital photo images is an unresolved problem yet. Textual information plays the key role in the understanding of photo images because so called “semantic gap” can not be overcome using pixel features only.

Any additional value of the image is due to its meta data – textual information that provides knowledge about the photo image: content of the image, author, when, where and how an image was created, etc. Much of the most valuable meta data, especially the descriptive and explanatory meta data, is added manually using the special software or by intelligent software (semi automatically or automatically) that is mostly based on the availability of labeled samples. Photographic equipment can not capture and record the meaning objects or events and yet more in the photos.

The content independent meta data (author’s name, date, location, etc.) are out of the scope of the work. We are interested in the content descriptive meta data, especially in the image annotations. Such information refers to the content semantics – real word objects, temporal events, emotions, etc.).

Descriptive information same as all other meta data can be stored separately and can be contained within the photo image file format itself. Meta data and digital image files are closely related components, and they have to be connected.

The multimedia meta data standard IPTC-IIM (the International Press Telecommunications Council–Information Interchange Model), often called “legacy” IPTC, enables users to insert and edit their own data (keywords, location, description, etc.) and stores it in an additional header section of the digital image file format.

The multimedia meta data standard introduced by Adobe XMP (Extensible Meta-data Platform) is used for storing meta data within an image file or in a separate text format file, often called “sidecar” file and allows the creation of custom meta data fields. XMP is a combination of XML and RDF. IPTC-IIM and XMP descriptive data can be integrated within many different image file types (PDF, PS, etc.), including digital photography files (JPG, JP2, TIFF, PNG). Each file format has distinct rules for storage of meta data within the file. An example, there are older formats (BMP, TGA, etc.) that does not support header information overall. Formats like JPEG 2000 or parts of TIFF and TGA define the concept of containers or boxes to store the information. When saving an image in a different format, one should be aware about its possibilities to preserve the descriptive information, as it can be lost when the file format is changed. The attached information is not a standard part of the image data and there is no uniform support for certain attributes in each image file format. Furthermore, the simply attached information creates additional payload when the image is transmitted.

Connection between descriptive information and digital image files can be modeled using XML that defines a platform-independent meta data exchange format and plays the substantial role on the Web. XML file format makes information about the image content portable and extendable. There are modern XML-based online annotation tools for outlining and labeling objects in the photo images. They can be used for photos search and retrieval, object detection and recognition research [7]–[9].

The information can be lost when descriptive information is stored separate from the object it describes. The problems of linking between the image file and the XML file arise from the variety of reasons: change of the image file name, limited data connection capabilities by the user etc. XML-based technology let us enter more text than embedded meta data technology and does not burden the volume of photo image file. Both technologies create an additional payload for transmission and storage of data.

Using information hiding methods [10] the embedded data becomes an inseparable part of the media and takes almost no additional storage space. However, only a few applications use these methods for annotation of the natural scene images.

Researchers present DCT-based watermarking for the color images in [11] and color spatial domain watermarking in [12]. Both methods segment the image in to the regions of interest and embed the content based information in each region. Authors of [12] suggest the described watermarking method can be applied for content-based indexing, retrieval and manipulation of digital images and image regions. Watermarking is one of application areas of information hiding but its requirements differ from methods of information embedding with annotative purpose [10].

Some of authors [13] have created the way for embedding annotations about separate objects into the digital photography content using the full capacity of the photo image object. Their algorithm operates in the frequency domain using the phase and magnitude components of a DFT. The descriptive data is embedded by modulating the

phase and the hierarchical visual-functional relations between the objects are embedded by modulation of magnitude in predefined frequency bands. The description and the hierarchical classification of the object is provided by the user. As distinct from our implementation last mentioned annotating technology based on DFT and requires additional input.

III. QUALITY EVALUATION OF THE NATURAL SCENE IMAGE

Quality of digital image depends on many factors: setting of camera, experience of photographer, quality of cameras and other. In this work we are not concerned on the initial quality of digital image, but focusing on keeping the result as close to the original, as possible.

From the statistical viewpoint, a natural image is a signal with certain statistical properties.

Whereas the ultimate receiver of annotated image is a human being, we have chosen the objective image quality assessment methods that are related with or guided by the human vision model in order to reflect human perception accurately. The subjective quality measurement Mean Opinion Score (MOS) is usually too inconvenient, time-consuming and expensive in practice, while standard pixel-based quantitative image quality metrics, like PSNR and MSE, are not directly related to human perception.

Our chosen the major and often-used full-reference image quality metrics. The structural similarity index (SSIM) and visual information fidelity (VIF) were compared with seven public image databases (totally 3832 test images), and the conclusion was made that SSIM-based and VIF are relatively better metrics [14].

HVS-based metrics SSIM [15], its multi scale modification MS-SSIM and the information content weighted structural similarity measure (IW-SSIM) [16] are interdependent. SSIM measures the similarity between two images. The perceptibility of image details depends on the sampling density of the image, the distance from the observer to the image plane and the perceptual capability of the observer’s visual system, thus varying the subjective evaluation of the images. SSIM because of single scale analysis is valid only for specific settings. Multi scaled MS-SSIM is able to evaluate image details at different resolutions. By combining information content weighting with multi scale analysis, IW-SSIM is defined as information content weighted measure.

Researchers believe that HVS models are the dual of the natural scene statistics (NSS) models. The visual stimulus emanating from the natural environment drove the evolution of the HVS. Many aspects of the HVS are modeled in the NSS description.

IV. IMPLEMENTATION OF ANNOTATION METHODOLOGY

It is possible to embed and retrieve the information into digital image, if the quality is not a concern [17]–[19] or loss of quality may be accepted in certain regions. The key point of the research is to verify the influence of information embedding into the image when introduced artifacts are unacceptable.

The embedding is performed into JPEG 2000 image using

unmodified scheme, presented in [19], when the annotation and the region of annotation (ROA) are not separated spatially.

The embedded structure of ROA contains information of the annotation, graphical finder patterns and error-correcting codes. The research uses 2D barcode symbology as a carrier, reassembling DataMatrix finder pattern and specially crafted message area. The initial consideration were made on using standard DataMatrix symbology, but as no public encoder and decoder were available, a simulated one has to be used.

The most promising candidate seems to be QR symbology. Minimal size of QR 2D barcode is 25×25 px and it needs additional 4 px margin, which could be reduced to 1 or 2 px, in the case of success. In this case it can transmit 20 to 47 characters, as defined by the level of error correction [17]. This size of barcode is too large to embed into lower levels of DWT decomposition, but more compact version, called "Micro QR" [18] could be used. Micro QR can transmit from 6 to 11 alphanumeric symbols in its smallest size of 13×13 px and having additional 2 px margin, and this is comparable to the 2D symbology used in the research.

Three publicly available images - "Lena", "Boats" and "Goldhill" were taken to evaluate the results. Every image is compressed using JPEG 2000 image compression and a single annotation of 50 symbols was embedded to the selected region (the principal part of the image). This is simulation of natural workflow, when the image is acquired, compressed and saved, but annotations are placed later on, in the next step.

Image quality is evaluated using IW-SIM and MS-SSIM criteria, similarity index of 0.75 is treated as "satisfactory". VIF quality criteria was not used due to the lack of research of what minimal index values are to be treated as "satisfactory".

V. EVALUATION OF EXPERIMENTAL RESULTS

As in [19], the embedding scheme is reliable enough to carry additional information, embedded into digital image. Image quality metrics for compression ratios up to 1:50 is not lower than 0.7 when compared to the original image and is around 0.9...0.98 when compared to the compressed one.

Typical behaviour of IW-SSIM and MS-SSIM values for annotated and compressed image are compared to the original image, are presented in Fig. 1. Typical values of IW-SSIM and MS-SSIM for annotated and compressed image are compared to the compressed image and are presented in Fig. 2. As it is seen from Fig. 2, the additional information embedded plays the minor role to the total loss of image quality.

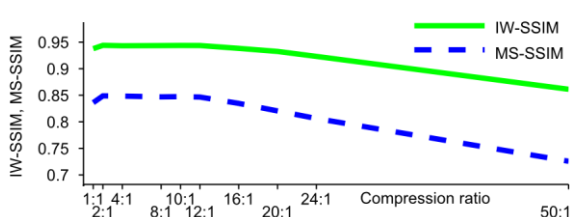


Fig. 1. Typical IW-SIM and MSSIM values for annotated and compressed image compared to the original image.

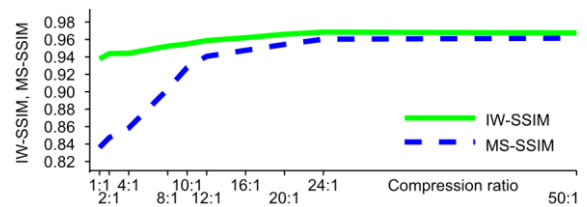


Fig. 2. Typical IW-SIM and MSSIM values for compressed and annotated image compared to the compressed image.

When comparing local distortions, the annotated region clearly reveals itself in the image (Fig. 3), and this is unacceptable for this image type. The pattern of the region suggests the distortion appear due to the intensive use of DWT decomposition tree, when the annotation is embedded to every subband of every decomposition level. This makes suggestion for using less intensive embedding schemes, like embedding to a single subband or a single decomposition level only. On the other hand, this will be the limiting factor for informational capacity.

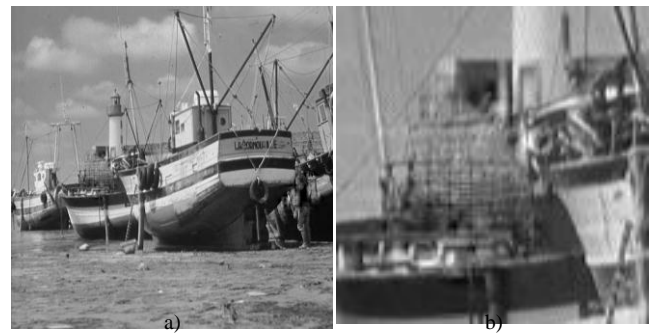


Fig. 3. Image "Boat" (a) and its annotation region (b). JPEG 2000 compression ratio 1:4.

No attention in the testing was paid to the retrieval rate of information, due to the poor visual performance of the annotation region. Preliminary testing shows about 70% of the embedded information can be retrieved. This positively correlates with permissible error rates of 2D barcode symbology, especially QR code.

VI. CONCLUSIONS

It is possible to embed and retrieve additional annotative information to the digital image.

The most significant image distortion is due to the intensive re-use of the same spatial area for annotation embedding.

Information retrieval rate can be adjusted using more robust carrier, e. g. using 2D barcode symbology with higher permissible error rate.

REFERENCES

- [1] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain, "Content-based Image Retrieval at the End of the Early Years", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1349–1380, 2000. [Online]. Available: <http://dx.doi.org/10.1109/34.895972>
- [2] T. Baba, T. Chen, "Object-driven Image Group Annotation", in *Proc. of the IEEE 17th International Conference on Image Processing*, 2010, pp. 2641–2644.
- [3] L.-Ch. Hsieh, W. H. Hsu, "Search-Based Automatic Image Annotation via Flickr Photos Using Tag Expansion", in *Proc. of the ICASSP*, 2010, pp. 2398–2401.

- [4] J. Y. Choi, W. D. Neve, K. N. Plataniotis, Y. M. Ro, "Collaborative Face Recognition for Improved Face Annotation in Personal Photo Collections Shared on Online Social Networks", *IEEE Transactions on Multimedia*, vol. 1, no. 13, pp. 14–28, 2011. [Online]. Available: <http://dx.doi.org/10.1109/TMM.2010.2087320>
- [5] L. Cao, J. Luo, H. Kautz, T. S. Huang, "Image Annotation within the Context of Personal Photo Collections Using Hierarchical Event and Scene Models", *IEEE Transactions on Multimedia*, vol. 2, no. 11, pp. 208–219, 2009. [Online]. Available: <http://dx.doi.org/10.1109/TMM.2008.2009693>
- [6] Z. Hua, X. J. Wang, Q. Liu, H. Lu, "Semantic Knowledge Extraction and Annotation for Web Images", in *Proc. of the 13th annual ACM International Conference on Multimedia*, 2005, pp. 467–470. [Online]. Available: <http://dx.doi.org/10.1145/1101149.1101253>
- [7] A. Torralba, R. Fergus, W. T. Freeman, "80 Million tiny Images: a large Data Set for Nonparametric Object and Scene Recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 30, pp. 1958–1970, 2008. [Online]. Available: <http://dx.doi.org/10.1109/TPAMI.2008.128>
- [8] B. Russell, A. Torralba, K. Murphy, W. Freeman, "LabelMe: a Database and Web-based Tool for Image Annotation", *International Journal of Computer Vision*, vol. 1, no. 77, pp. 157–173, 2008. [Online]. Available: <http://dx.doi.org/10.1007/s11263-007-0090-8>
- [9] H. Astudillo, V. Codocedo, G. Canals, D. Torres, A. Diaz, A. Napoli, A. Gomes, M. Pimentel, "Combining Knowledge Discovery, Ontologies, Annotations, and Semantic Wikis", Author manuscript, published in „Webmedia Minicourse Book SBC“, 2009.
- [10] G. Kazakevičiūtė, E. Januškevičius, R. Rosenbaum, H. Schumann, "Self Annotated Raster Image", *Information Technology and Control*, vol. 2, no. 35, pp. 106–116, 2006.
- [11] G. Lo-varco, W. Puech, M. Dumas, "Content Based Watermarking for Securing Color Images", *Journal of Imaging Science and Technology*, vol. 5, no. 49, pp. 464–473, 2005.
- [12] S. V. Mezaris, N. V. Boulgouris, I. Kompatsiaris, D. Simitopoulos, M. G. Strintzis, "Segmentation and Content-based Watermarking for Image Indexing and Retrieval", *EURASIP Journal on Applied Signal Processing*, vol. 4, no. 2002, pp. 418–431, 2002.
- [13] C. Vielhauer, M. Schott, C. Kraetzer, J. Dittmann, "Nested Object Watermarking: Transparency and Capacity Evaluation", in *Proc. of the SPIE conference at the Security, Steganography, and Watermarking of Multimedia Contents X*, 2008.
- [14] W. Lin, J. Kuo, "Perceptual Visual Quality Metrics: a Survey", *Journal of Visual Communication and Image Representation*, vol. 4, no. 22, pp. 297–312, 2011. [Online]. Available: <http://dx.doi.org/10.1016/j.jvcir.2011.01.005>
- [15] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, "Image Quality Assessment: from Error Visibility to Structural Similarity", *IEEE Transactions on Image Processing*, vol. 4, no. 13, pp. 600–612, 2004. [Online]. Available: <http://dx.doi.org/10.1109/TIP.2003.819861>
- [16] Z. Wang, Q. Li, "Information Content Weighting for Perceptual Image Quality Assessment", *IEEE Transactions on Image Processing*, vol. 5, no. 20, pp. 1185–1198, 2011. [Online]. Available: <http://dx.doi.org/10.1109/TIP.2010.2092435>
- [17] *QR code introduction*, Symbol version, 1 to 10, Denso Wave Inc., 2012. [Online]. Available: <http://www.denso-wave.com/qrcode/vertable1-e.html>
- [18] *QR code introduction*, Symbol version, Micro QR code, Denso Wave Inc., 2012. [Online]. Available: <http://www.denso-wave.com/qrcode/microqr-e.html>
- [19] G. Kazakevičiūtė-Januškevičienė, E. Januškevičius, "A new Approach for Raster Images Annotation", in *Proc. of the 17th International Conference on Information and Software Technologies, IT 2011*, 2011, pp. 182–189.