Metadata Representation for Semantic-Based Multimedia Security and Management

Yuriy Rytsar, Sviatoslav Voloshynovskiy and Thierry Pun

Computer Science Department, Stochastic Image Processing Group,
University of Geneva, 24 rue du General Dufour,
CH-1211 Geneva 4, Switzerland
{Yuriy.Rytsar, Svyatoslav.Voloshynovskyy,
Thierry.Pun}@cui.unige.ch
http://sip.unige.ch/index.html

Abstract. In this paper a novel approach to metadata representation for semantic-based multimedia security and management is proposed. This approach is based on semantically meaningful object segmentation and digital data hiding to facilitate image indexing as well as identification for dedicated image databases. An image is considered as a collection of the regions that correspond to objects, where these objects are associated with some hidden metadata describing some of their features. The metadata is hidden inside the raw images themselves for security and "portability" reasons. The embedding of hidden metadata allows moving an image from one database to another, as well as the insertion/cropping of objects from one image to another, while still preserving the associated descriptions. The perceptual invisibility of the integrated metadata confirms the high performance of this proposed object-based hidden metadata representation.

1 Introduction

Due to enormous increase in number of Web-based distributed environments, applications and public networks, an efficient management of digital multimedia database systems became over the last decade one of the important and challenging problems for modern multimedia applications. Taking into account the practical difficulties in controlling and verifying of existing audio/visual communications in distributed and public networks, one can imagine additional ambiguity and insecurity in extracting and managing multimedia information. Therefore, one of the possible solutions to tackle the above problem is to use the metadata. Most of Web-based applications strongly rely on metadata features for multimedia management, security and communications. Thus, in multimedia management the metadata is usually applied to facilitate proper cataloging and indexing of large amounts of visual information for effective browsing and search.

To provide complete information about the ownership of multimedia source or to describe the copyright or/and licensing information, the metadata can be easily utilized as well. At the same time, security issues around semantic features, annotations, metadata, object "transportability" still need to be addressed. For instance, even the simple change of file name can trick the Google image search engine and instead of the targeted data can display prohibited or age censored content. The headers of the JPEG/JPEG2K files containing metadata description can be modified exactly in the same way while the object cropping/insertion from one content to another can change semantic features, educational statements, historical or even criminal evidences of the retrieved content. The quality-of-service (QoS) control in distributed networks and particularly in heterogeneous or time-varying networks can also be provided by metadata, describing information about the channel abilities and properties. Therefore, the usage of metadata becomes an obvious and effective way to protect, manage and distribute multimedia information for current multimedia applications.

To introduce our concept of integrated metadata in different applications for multimedia management, security and communications we first define some general concepts and meanings. In the most general sense, the term of metadata means “data about data”. However, in the context of images this term refers to image metadata - “data associated with an image” or “data about an image”. In other words, the image metadata can be considered as all non-pixel data associated with an image or as tags, which are stored within image files [1]. In this paper, we do not follow the classical separation principle where metadata is treated independently from the raw pixel data. On the contrary, we propose to “merge” them together in a simple and elegant way using data hiding technologies.

Image metadata or data describing digital images must contain information about the image, when it was taken, by whom, with what equipment (digital or web-camera, scanner, mobile phone), copy/distribution restrictions, etc. The metadata can be generated after the image has been created to allow an event description or scene annotations. Almost all digital cameras register information about time and date of captured picture, image dimensions, etc. Since metadata becomes an integral part of the image it is necessary to provide appropriate security features against their intentional and unintentional modifications. In general even simple modification of image format will unavoidably lead to the change of metadata content (for example, from JPEG to PNG format). Unfortunately, the limited capabilities of current Web-based environment and format compatibility do not provide an adequate level of metadata attribute extraction and its sufficient protection. Therefore, new models and methods are needed for metadata protection and security while preserving efficiency.

In this paper, we present one of the possible ways to solve the above problems of metadata efficiency and security by the embedding of image annotations inside the raw image data. Thus, in digital database the images will be indexed by their own visual content instead of being annotated by text-based keywords and the metadata will be completely hidden for non-authorized access. Moreover, this approach will provide more security for image information as well as the image “portability” from one digital database to another without additional associated descriptions.

This paper is organized as follows. In Section 2 we describe interactive image segmentation for content-based image extraction. The object-based metadata hiding approach is described in Section 3. In Section 4 we demonstrate experimental results of region-based segmentation with metadata hiding. Section 5 provides some concluding remarks.

2 Interactive Segmentation for Content-Based Image Extraction

The main difficulty of existing image and video database indexing and searching systems [2-5] is that they cannot automatically describe multimedia content using both low-level (color, shape, texture) and semantic (objects, people, places) descriptors [6]. Usually, the researchers utilize various complex visual features like color distribution, object shapes, textured surface or simple textual descriptions like annotations, keywords, queries or captions to search in large databases. Besides, the usage of the semantic features requires a good correspondence between the description and the content of multimedia data. Another problem arises when an object is cropped or inserted from one image into another image. In this case the integrated descriptions will be completely lost due to the fact that metadata is separated from image data.

The other difficulty is that the same image content may be interpreted by different people in completely different ways. Subjective human descriptions may cause ambiguity and mismatches in the extraction process and in the management of multimedia information. The relevance of visual content significantly depends on the subjectivity of both a database provider and a user. Although a lot of study and publications are dedicated to this area and many retrieving systems are already developed, there are still many open research and commercial issues to be solved and applied in practice [7-10].

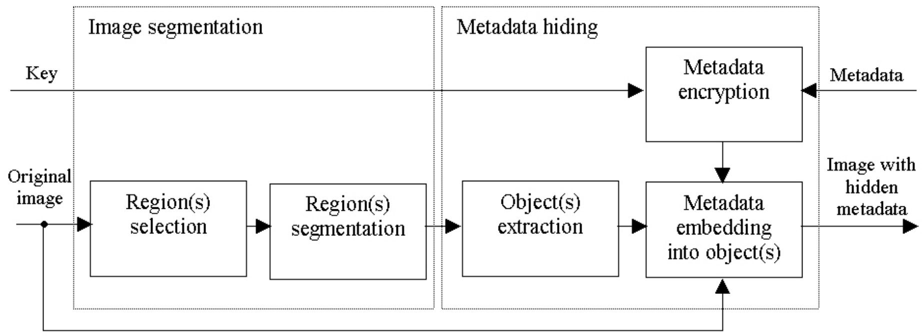


Fig. 1. Block diagram of metadata embedding

We propose the following content object-based retrieval approach as a possible solution to the above mentioned problems. It is based on a human-computer interface for semantics extraction and object-based metadata integration by using digital data hiding technologies. Instead of being annotated by text-based keywords, the images may be labeled by their own semantic content, which is hidden into every image object. Since each homogeneous region in the image has common features (brightness, color distribution, textured surface) it is possible to exploit this fact for an object-based extraction by performing semantic object segmentation. Different objects can be semantically distinguished regarding their own features and easily identified by a search engine. The problem of classical unsupervised segmentation is its inability to capture semantic relationships in real world images. That is why human assistance and feedback are needed to guide the process of semantically meaningful segmentation in order to make an adequate correspondence of each object of the image to its integrated metadata content.

In this paper we do not focus on specific segmentation algorithms (for example, [11-14]), we rather refer to the fact that successful multimedia retrieval critically depends on the chosen segmentation technique based on region or object content. The successful segmentation of the image is heavily dependent on the criteria used for the merging of pixels based on the similarity of their features, and on the reliability with which these features are extracted. An unsupervised stochastic segmentation is used as a first iteration and the result is displayed for the user. The user then defines where the region of interest is, and the chosen segmentation algorithm performs partitioning into meaningful objects. The human-computer interface is organized in such a way that the user can easily add/remove some objects or parts of objects from previous stages, select objects of interest, or even merge some objects that are classified as distinct according to their statistical properties but in fact represent parts of the same semantically connected object.

Hence, the image is not considered as a set of pixels but rather as a set of annotated regions that correspond to objects or parts of objects, where these objects are associated with some hidden descriptive text about their features. Each selected object of the image carries its own embedded description that makes it self-containing and formally independent from the particular image format used for storage in image databases. Therefore, the proposed approach of metadata representation can easily provide "portability" of each object from one image to another, as well as makes resistant to object insertion or/and cropping, which are usually used in multimedia processing and management. Moreover, the image can be moved from one database to another without any associated descriptions because of the self-containing features of objects. One can consider this feature as a joint distribution of visual and textual description (image and text).

The block diagram of the proposed approach for metadata embedding into image objects is shown in Figure 1. First, the user selects region(s) of interest on a given image. Region-based segmentation is then applied for pre-selected region(s) in order to extract object(s) of interest for their further description. Secondly, the user embeds the previously encrypted metadata (text description or short annotation) inside the selected object(s) based on robust digital data hiding technique and personalized secret key.

3 Object-Based Metadata Hiding

The second part of the proposed approach is based on the hidden indexing/labeling of image objects by using robust digital data hiding technique. Here, we are focused neither on the process of metadata encryption and hiding (see more in [15]) nor on the robustness of data hiding algorithms against some attacks; this is out of the scope of this paper. We rather discuss the possible ways where and how to integrate the metadata reliably into the image and then, how to extract it without errors. We intend to embed the semantically segmented map as well as the user assigned description into the body of each object as the robust watermark. Thus, the image is considered as a set of "smart objects" divided by boundaries that can perform:

- self-indexing;
- self-authentication;
- self-synchronization;
- self-extraction;
- self-tamper proofing.

In other words, the image contains all necessary information about itself including descriptions of the object bodies, of their mutual allocation in images, as well as complementary hidden metadata. Moreover, no additional header, attachment, tag or extra metadata are needed for further image indexing and identification.

It should be noted that the metadata embedding/extraction could be based on symmetric or asymmetric protocol en/decryption. The *symmetric protocol* means that only the authorized party can embed, modify and retrieve the hidden information by using a private secret key. The *asymmetric protocol* means that only the authorized party can embed and modify the hidden metadata by using a private secret key, and that the rest of users can only extract it by using a public key. In both cases the additional security level of metadata is provided. Consider one possible example of the proposed system for telemedicine application: for confidentiality reasons it is based on a symmetric protocol. Each region of interest has its own label with hidden description connected to the extracted object, and the private secrete key is used for information embedding. For example, a physician working with the MRI of a patient after having used the interactive segmentation tool for the detection of possible tumor or clot of blood will be able to insert the necessary information directly into the image or into some parts of this image. Even if the patient or any unauthorized party recovers his MRI data he will be unable to extract this description without secret key. However, the patient can give this image to another physician, who will be able to directly extract this invisible hidden information using the same segmentation tool and the proper private key. Therefore, time and money will be saved as well as security and confidentiality of this protocol will be provided.

As a possible extension, the semantic content of the object body can be replaced by other hidden multimodal media data (see more in [15]). For example, each object may have its own audio/musical context and this feature may be used for blind people (special interfaces, web-browsers). Another possible extension would consist in encryption or steganographical applications, meaning the hiding of small amounts of textual or visual raw-data into selected “smart objects”. Besides, the raw-data about the contour of the object body can be additionally embedded into the same object. In other words, the image has more information about itself and this information is perceptually invisible and is not integrated inside image headers.

The block diagram for metadata extraction from the image objects is shown in Figure 2. First, the user selects the region(s) of interest on a given image. Region-based segmentation is then applied for the pre-selected region(s) in order to extract object(s) of interest to obtain their descriptions. Secondly, the user retrieves the metadata from the extracted object(s) based on the secret key.

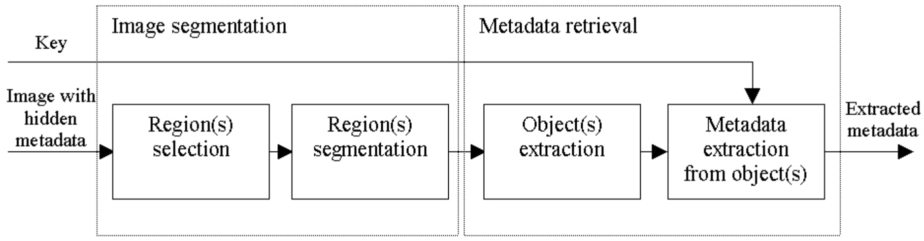


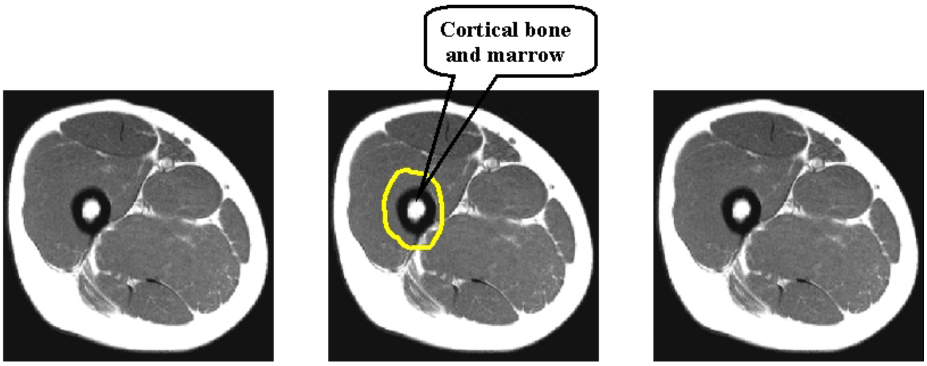
Fig. 2. Block diagram of metadata extraction

Obviously, this visual and textual representation of the image needs additional efforts to keep all information more secure and robust to possible modifications and transformations. Besides, the synchronization of the resulting segmentation process is required for both metadata embedding and extraction. After reliable and successive application of the segmentation technique only an authorized party can extract object(s) of the image to retrieve the hidden metadata bits associated with the given object(s) by using a secret private key. The invariance to different types of distortions can be achieved by providing additional embedding of hidden information into image objects, robust to such changes. Therefore, for robustness reasons the object descriptions can be additionally encoded by error correction codes (ECC) (Turbo code [16] or LDPC code [17], for example) depending on the applications.

In order to avoid possible multiple insertion of contradictory metadata into the same image object we consider several scenarios, which can be applied for the metadata embedding. As the first scenario, we propose to perform preliminary detection before the insertion of metadata information into selected image object. If the preliminary metadata extraction is successful no additional information will be embedded into the object. Another possible scenario is complete metadata replacement, where the previously embedded information related to the object is completely replaced by new data by the user request. In the case, when it is necessary to add complementary information to the embedded metadata we propose to check the “free space” for the location of an additional raw pixel data. If this space is enough to perform the additional embedding, one can also replace completely the metadata information or leave it unchanged depending on the scenario.

4 Application Scenarios

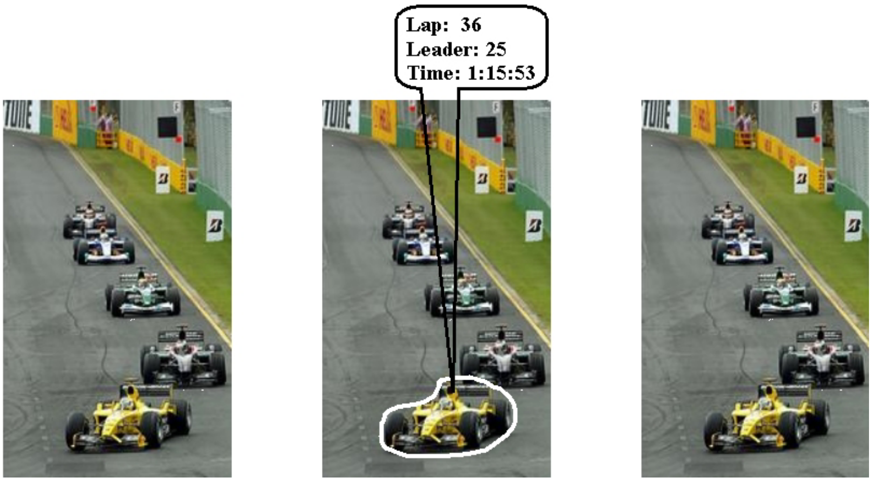
The performance of the presented technique was tested for embedding and extraction of the metadata for image objects on different grayscale and color images. Here are considered two application scenarios of the proposed approach for visual annotations of the images and for image indexing and search in databases. In Figure 3 the experimental results of metadata hiding for medical MRI image of axial slice of the human femur are shown. The region of interest is defined and marked by using a human-computer interface. The corresponding textual information about the selected image object (cortical bone and marrow) is encrypted based on the secret key, and is then embedded into marked image regions. One can find that the image quality with embedded metadata is not perceptually degraded. The PSNR value is equal to 38 dB.



(a) (b) (c)

Fig. 3. Integrated metadata for medical MRI image: (a) original medical MRI image of axial slice of the human femur; (b) visually annotated image with marked region of interest; (c) the resulting image with hidden visual annotation

For the second application scenario, we used several images from an official Formula 1 car racing website [18] (Figure 4a, 5a). We performed local unsupervised region-based segmentation in order to define regions of interest for the original image. After, we selected and merged together small segmented objects into one semantically meaningful object by using the human-computer interface (Figure 4b). Then, the marked object was labeled by a corresponding textual description and integrated into the object by using a symmetric encryption procedure (Figure 4c). This metadata was used for further indexing and search for content-based databases.



(a) (b) (c)

Fig. 4. Integrated metadata for the car racing image: (a) original image; (b) image with marked object; (c) the resulting image with hidden description

We copied and inserted the marked object (yellow car) into another test image (see Figure 5b). Since the metadata was integrated and hidden into image objects, the description information was simultaneously "transported" with this object to others images, while image quality was completely preserved (see Figure 5c). During the retrieval stage all hidden metadata was successfully extracted from this test image based on the same secret key.

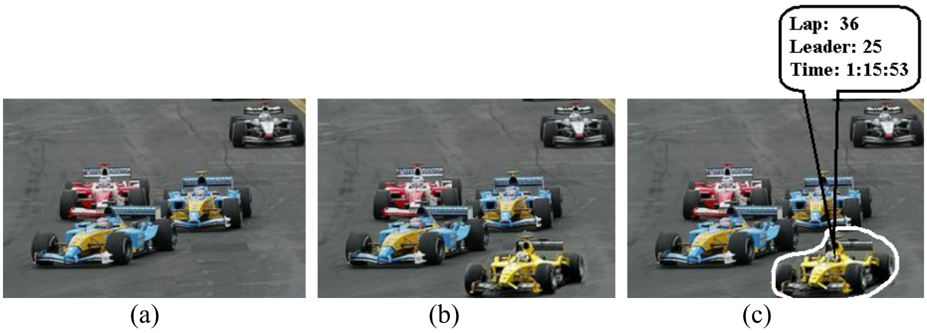


Fig. 5. Extracted hidden metadata for the target image: (a) target image; (b) image with inserted object; (c) the resulting image with extracted hidden description

Since the identical object is present in both images, the only possible way to distinguish and retrieve them from an image database is based on the proper hidden metadata of this object, but not an the visual data itself. We mean the corresponding textual description such as general information about object, its identification number, date and time of image creation, ownership, spatial size of given object, short information about the image to which this object belongs.

Here we are not focused on copyright protection, but we rather demonstrate the ease with which the object information can be kept reliably and properly. Obviously, the retrieved metadata from those images can be used to prove the origin of the photographs, as well as to outline the modifications that have been performed on them (see more details about the tamper proofing and authentication techniques in [19, 20]).

It should be noticed that for both application scenarios the image quality with embedded metadata is not perceptually degraded. Besides, on the second example we demonstrated the "portability" features of our technique where one image object can be easily inserted into another image without visible degradation and loss of embedded information.

5 Conclusion

In this paper a hidden metadata representation approach has been proposed for content management and reliable communications. The presented technique applies interactive object-based segmentation for metadata hiding into extracted objects of the image in order to provide additional protection and security. Moreover, the embedded metadata is independent from the particular format used for storage in an image database, and is robust to any format changes as well as to insertion/cropping image transformations. Besides, the perceptual invisibility of the integrated metadata confirms the high performance of proposed approach.

Acknowledgment

This work was partially supported by the Swiss National Center of Competence IM2 – Interactive Multimedia Information Management.

References

1. Milch, J.: Eastman Kodak Company. Hints for the Creation and Usage of Picture Metadata (2000)
2. Smith, J.R., Chang, S.-F.: VisualSEEK: A fully automated content-based image query system. Proc. ACM Multimedia'96 (1996)
3. Ogle, V.E., Stonebraker, M.: Chabot: Retrieval from relational database of images. Computer, Vol. 28. 9 (1995) 40-48
4. Müller, H., Müller, W., Marchand-Maillet, S., Squire, D. McG., Pun, T.: A Web-Based Evaluation System for Content-Based Image Retrieval. Proceedings of the Multimedia Workshop on Multimedia Information Retrieval, Ottawa, Canada (2001)
5. Gevers, T., Smeulders, A.W.M.: Pictoseek: combining color and shape invariant features for image retrieval. IEEE Trans. on Image Processing, Vol. 9. 1 (2000) 102-119
6. Xu, Y., Saber, E., Tekalp, A.M.: Object Segmentation and Labeling by Learning from Examples. IEEE Trans. on Image Processing, Vol. 12. 6 (2003) 627-638
7. Damiani, E., De Capitani di Vimercati, S., Fernández-Medina, E., Samarati, P.: An Access Control System for SVG Documents. Proc. of the Sixteenth Annual IFIP WG 11.3 Working Conference on Data and Application Security, King's College, University of Cambridge, UK (2002)
8. Müller, H., Müller, W., Squire, D. McG., Marchand-Maillet, S., Pun, T.: Performance Evaluation in Content-Based Image Retrieval: Overview and Proposals. Pattern Recognition Letters, Special Issue on Image and Video Indexing. (eds.): H. Bunke and X. Jiang, Vol. 22. 5 (2001) 593-601
9. Rui, Y., Huang, T. S., Chang, S.-F.: Image retrieval: Past present and future. J. of Visual Communication and Image Representation, Vol. 10. (1999) 1-23

10. Boulgouris, N.V., Kompatsiaris, I., Mezaris, V., Simitopoulos, D., Strintzis, M.G.: Segmentation and Content-based Watermarking for Color Image and Image Region Indexing and Retrieval, EURASIP Journal on Applied Signal Processing, 4 (2002) 420-433
11. Zhang, Y.J.: Evaluation and Comparison of Different Segmentation Algorithms. Pattern Recognition Letters, Vol. 18. 10 (1997) 963-974
12. Pal, N.R., Pal, S.K.: A review on image segmentation techniques. Pattern Recognition, Vol. 26. 9 (1994) 1277-1294
13. Kurugöllü, F., Sankur, B., Harmanci, E.: Multiband Image Segmentation Using Histogram Multithresholding and Fusion. J. of Image and Vision Computing, Vol. 19. 13 (2001) 915-928
14. Duygulu, P., Vural, F.: Multi-Level Image Segmentation and Object Representation for Content Based Image Retrieval. SPIE Electronic Imaging 2001, Storage and Retrieval for Media Databases. San Jose, CA (2001)
15. Rytsar, Y., Voloshynovskiy, S., Ehrler F., Pun, T.: Interactive Segmentation with Hidden Object-Based Annotations: Toward Smart Media. SPIE Electronic Imaging 2004, Storage and Retrieval Methods and Applications for Multimedia. San Jose, CA (2004) *(accepted)*
16. http://www331.jpl.nasa.gov/public/JPLtcodes.html
17. Gallager, R.G.: Low Density Parity Check Codes. Monograph, M.I.T. Press (1963) (http://www.inference.phy.cam.ac.uk/mackay/gallager/papers/ldpc.pdf)
18. Australian Grand Prix 2003 postcards: http://www.formula1.com/gallery/images/2/Sunday/1949.html
19. Voloshynovskiy, S., Deguillaume, F., Koval O., Pun, T.: Robust digital watermarking with channel state estimation: part II Applied robust watermarking. Signal Processing (2003) *(submitted)*
20. Deguillaume, F., Voloshynovskiy, S., Pun, T.: Secure hybrid robust watermarking resistant against tampering and copy-attack. Signal Processing (2003) *(submitted)*