International Journal of Advanced Computer Technology (IJACT)

ISSN:2319-7900

AUTOMATIC INJECTION OF DEEPER SEMANTICS FOR EFFECTIVE IMAGE SEARCH

Clement H.C. Leung, United International College; Yuanxi Li, Hong Kong Baptist University

Abstract

With the number of social networking and photograph sharing sites growing at an unprecedented pace, image search queries has become a popular and frequent activity for Internet users, and this has become increasingly challenging due to both the large number of sites dedicated to image hosting as well as the number of conventional sites incorporating significant image elements. Such a trend is expected to further accelerate with the diverse types of devices that are able to capture digital images, and the wide variety of software that are able to enhance, edit and create them. As the number of Web images continues to increase, searching them efficiently and semantically becomes an important challenge. To meet this challenge, we present a comprehensive fully automated approach based on the analysis of image metadata in conjunction with image analysis techniques. In consequence, our model is able to automatically fill up the Structured Annotation fields in the MPEG-7 Description Standard which previously could only be performed manually. We envisage that such standard semantic fields may be routinely incorporated into image files similar to current metadata fields such as timestamps to take image retrieval to a deeper semantic level. Our system is evaluated quantitatively using more than 100,000 web images and around 1,000,000 tags, and our approach is able to yield highly promising results.

Introduction and Related Work

A convenient mechanism for searching images has increasingly become an integral part of our daily lives. For a small number of images, we may tag them one-by-one manually and search them afterwards via these tags. However, tagging images manually is very time consuming and involves large amounts of manpower. With a growing number of images to deal with, it is becoming essential and that this be done in an automatic way. Furthermore, with human tagging, it is sometimes difficult to define the strategy to ensure consistency of tagging and annotation for quality image search. Search and retrieval models are vital parts of image management, and are increasingly receiving attention with the growing volumes of web image uploads and the extensive use of image libraries. In this paper, we present a fully automated approach to inject deeper level semantics into images to enable their effective search and discovery. In the same way as timestamp and other metadata are already associated with images, using this method, standard semantic fields may be similarly and routinely attached to images to enable their retrieval based on their semantic contents.

Although some research studies [7, 10, 12, 15, 16, 20, 21, 28, 30, 39] aim to relate low-level image content to higher-level concept, these are limited to few isolated words only. Current image search systems, such as Yahoo! and Google, use some surrounding text or caption description provided by the users as the basis of keyword searches. These techniques have limited effectiveness at present and they need further development and refinement. While there are many different approaches and frameworks proposed for indexing and retrieval systems, they largely fall into two categories, namely, concept-based methods and content-based methods. The former focuses on retrieval by objects and high-level concepts [2, 4, 11, 14, 22, 23, 26, 29, 33, 35], while the latter focuses on searching large image repositories according to the low-level content of the image. With content-based methods, they aim at accessing the knowledge embedded in images by extracting low-level features and capturing image similarity and some specific characteristics of images [1, 3, 3, 9, 37, 38, 41].

The extraction of semantic concept from the metadata of images have been become one of the most active research areas. The works [5, 6, 13, 18, 31] conduct statistical analysis on the data fields of metadata to discriminate and classify scenes of images. In [18, 31], researchers make use of image metadata to address the problems of classifying images into mutually exclusive classes and use "scene mode" of the acquisition device to provide scene classifications. Although their contributions focus on binary classifications of images, these studies offer both strategic directions for future works and provide useful steps towards progress in this area.

In this article, we develop a comprehensive method in building an image search system based on automatic semantic tagging. Starting from the Automatic Semantic Annotation (ASA) [40] we enrich the semantics of images by using knowledge-based expansion and contextual feature-based expansion.

To facilitate quality image search, MPEG is responsible for the development of digital multimedia standards. Among the various standards, the most relevant to multimedia database management is MPEG-7 which provides description tools for facilitating searches and identifying objects. The MPEG-7 Structured Annotations Datatype is a comprehensive and potent standard [27, 33, 36] in terms of semantic richness for multimedia information. It can also be applied effectively to the image annotation domain [1, 32, 34]. It allows a simple structured description for audio visual mate-

ISSN:2319-7900

rials by using a number of fields, such as What Object, When, Where, What Action, Who, Why. In this paper, we shall make use of our approach to automatically furnish values to some of the MPEG-7 Structured Annotation fields without manual involvement, which may then be routinely attached to images as semantic fields similar to other non-content-based metadata currently in use.

Exploiting the Relationship between Metadata and Semantic Content

In the photographic world, many images may be broken down to several basic scenes and categories [19, 40], such as nature, wildlife, portrait, landscape and sports. A landscape scene comprises the visible features of an area of land, including physical elements such as landforms, living elements of flora and fauna, abstract elements such as lighting and weather conditions. Landscape photography is a normal approach to ensure as many objects are in focus as possible, which commonly adopts a small aperture setting. Sports photography corresponds to the genre of photography that covers all types of sports. The equipment used by a professional photographer usually includes a fast telephoto lens and a camera that has an extremely fast exposure time that can rapidly take photos.

Therefore, there are definite relationships exist between the scenes of image and image acquisition parameters. Fig. 1 lists out some typical image categories and scene images where scene images are subset of the corresponding image categories. The clustering behavior of images with respect to the image acquisition parameters is illustrated in Fig. 2.

The primary idea of our ASA approach is to group images based on embedded image-capture metadata and camera acquisition properties. As there are relationships between the type of scenes and image acquisition parameters, from parameters embedded in images, we may extract the intended scenes of images or semantic concepts from images.

We develop our automated annotation system by using an image database which consists of a collection of images obtained from a photograph album over the Web at random. All images in the database are metadata-embedded and stored in JPEG format. Since those images have no tags associated with them, we manually label all images with semantic concept (i.e. the scene of images) before arbitrarily dividing this image database into two groups, a training set and a test set. We use the training set to build our annotation model while the test set to measure its performance in terms of its error rate.

We construct our algorithm first from the training set by using decision tree technique, then measure its performance by the test set. After structured learning procedures with hundreds of testings, the best rules are obtained [40]. Some sample rules are given in Fig. 3 and a comprehensive listing of the rules are given in [40].

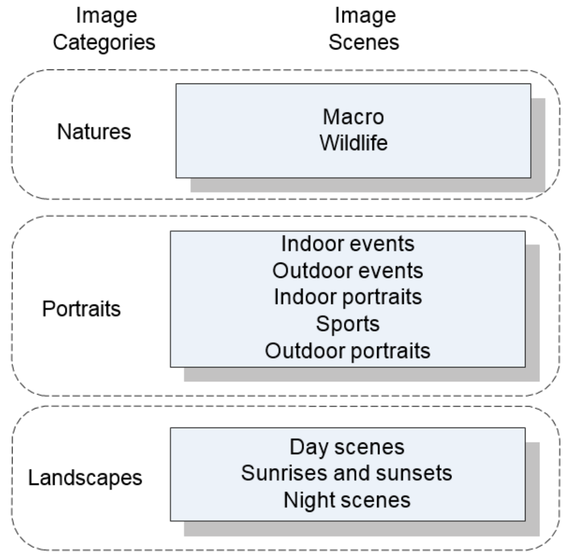

Figure 1. Categories and scenes of images

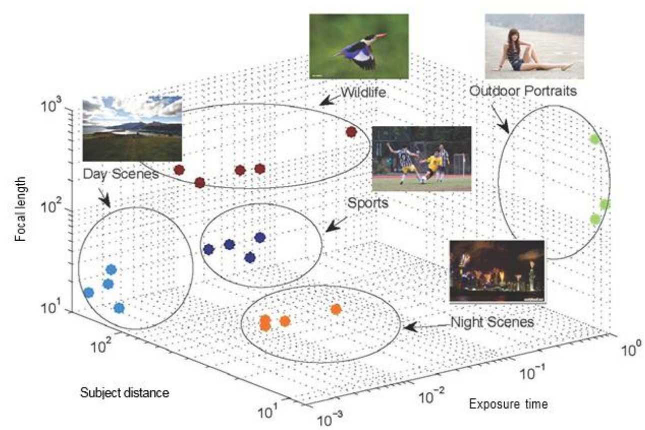

Figure 2. Image distribution in 3D space

$$
\begin{aligned}
&\forall i \in I, [(f_i \le 5.6) \wedge (5 < d_i \le 8)] \wedge \{[(t_i \le 0.00625) \\
&\wedge (L_i \le 30)] \vee [(30 < L_i \le 182) \wedge (ISO_i \le 250)] \\
&\vee (L_i > 182) \vee (t_i \le 0.003125)\} \Rightarrow i \in S_{op} \\
&\forall i \in I, (f_i > 5.6) \wedge (L_i \le 25) \wedge (5 < d_i \le 8) \\
&\wedge (t_i > 0.003125) \Rightarrow i \in S_{oe} \\
&\forall i \in I, (f_i > 5.6) \wedge (0.003125 < t_i \le 0.011111) \\
&\wedge (5 < d_i \le 8) \wedge (L_i > 25) \Rightarrow i \in S_{ip} \\
&\forall i \in I, (5 < d_i \le 8) \wedge \{(f_i \le 5.6) \wedge \{[(L_i \le 30) \\
&\wedge (t_i > 0.00625)] \vee [(ISO_i > 250) \\
&\wedge (30 < L_i \le 182)]\}\} \vee [(h_i = 1) \wedge (f_i > 5.6) \\
&\wedge (L_i > 25) \wedge (t_i < 0.011111)] \Rightarrow i \in S_{ie} \\
&\forall i \in I, (d_i > 10) \wedge (150 < L_i \le 400) \\
&\wedge (t_i \le 0.005) \Rightarrow i \in S_s
\end{aligned}
$$

Figure 3. Sample rules of Automatic Semantic Annotation

ISSN:2319-7900

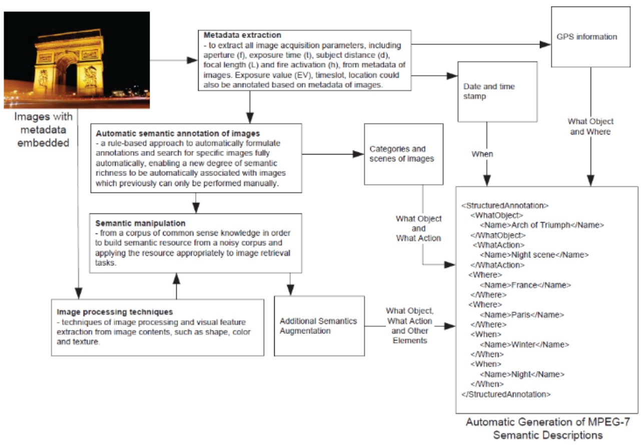

Figure 4. System structure

Annotation Enrichment and Extension

We annotate images with predefined semantic concepts in conjunction with methods and techniques of image processing, visual feature extraction and semantic concept manipulation. The system structure is shown in Fig. 4. The first step is image meta-data extraction followed by using decision trees and rule induction to develop a rule-based mechanism to formulate automatic annotations algorithms. The second step is the processing of image content, such as color, shape, texture and face detection. The third step is semantic concept manipulation which works in conjunction with common sense knowledge in order to build semantic resources from a noisy corpus and apply those to the image retrieval task. Subsequently, semantics augmentation and enrichment are carried out in conjunction with external data and additional algorithms, such as the weather log, news data or face recognition. Lastly, from these, we automatically furnish values to the applicable fields within the Structured Annotation Description of MPEG-7.

Our earlier work [40] covered the first two steps while this paper focuses on the third and fourth steps. A major difference between this paper and our prior work is both knowledge-based expansion and contextual feature-based expansion are incorporated. Furthermore, since MPEG-7 does not specify how such descriptors are obtained, this paper provides an automatic mechanism for generating of MPEG-7 Semantic Descriptions which previously could only be performed manually.

To extend our original ASA model, we enrich the annotation through the image contents. Algorithms based on color tone are widely used due to compactness of the calculations and information expression. Histogram is mainly used for color tone information and global color distribution is an important tool for the retrieval of images and video from multimedia databases. We convert all images in our training set into global color distribution format and plots into a scatter graph. There are only a small amount of outliers are distributed in its surrounding. This is because color image are tends to be centralized unless a specific concept of the image are intended to be present, such as "night scenes by the sea", "endless blue sky" or "sunrises and sunsets" which offers rich semantic meaning.

Here, we adopt edge detection algorithms with near-circular Gaussian-based image derivative operators [25] to extract high-level concepts from image concepts to match the similar images. The edge detection operators are based on first and second derivative approximations, corresponding to a first directional derivative $\partial u/\partial b \equiv b \cdot \nabla u$ and a second directional derivative $-\nabla \cdot (B\nabla u)$ and are defined by the functionals [8]

$$E_i^{\delta}(U) = \int_{\Omega} b_i \cdot \nabla u \zeta_i^{\delta} d\Omega \tag{1}$$

and

$$Z_i^{\delta}(u) = \int_{\Omega} \nabla u \cdot (B_i \nabla \zeta_i^{\delta}) d\Omega \tag{2}$$

where $B = b\, b^T$ and $b = (cos\theta, sin\theta)$ is the unit direction. The special case of the Laplacian operator is represented by Z_i^{δ} with B taken to be the identity matrix [8]. Firstly, we annotate night scenes based on the original ASA approach. We manually annotate one downtown image from the night scenes image set and then process all images by Gaussian kernel with $\sigma = 0.6$. By covering all image set, images were evaluated against the target image according to their similarity measures.

Methods using color tone is robust with respect to object movement, rotation and changes like distortion. Color is defined by the three characteristics: hue, saturation and value. HSV is an expression of color tones that can be sensed by humans using these characteristics.

An image histogram gives the probability function of the image intensities. This is extended for color images to capture the joint probabilities of the intensities of the three color channels. Histogram search characterizes an image by its color distribution. We adopt several color histogram distance algorithms [17, 30] which include the Bhattacharya Distance (BD),

$$d_{Bhattacharya}(p,q) = \sqrt{1 - \sum_{u=1}^{n} \sqrt{q_u p_u(m)}} \tag{3}$$

Chi-squared Distance (CD),

ISSN:2319-7900

$$d_{\chi^2} = \sum_{u=1}^{n} \frac{(q_u - p_u(m))^2}{(q_u + p_u(m))} \tag{4}$$

and Euclidean Distance (ED),

$$d_{euclidean} = \sqrt{\sum_{u=1}^{n} (p_u - q_u)^2} \tag{5}$$

where *m* is the center of the image region, *n* is the number of bins in the distribution, and q_u and p_u are the weighted histograms of the model and candidate respectively. Examples of these are given in Fig. 5.

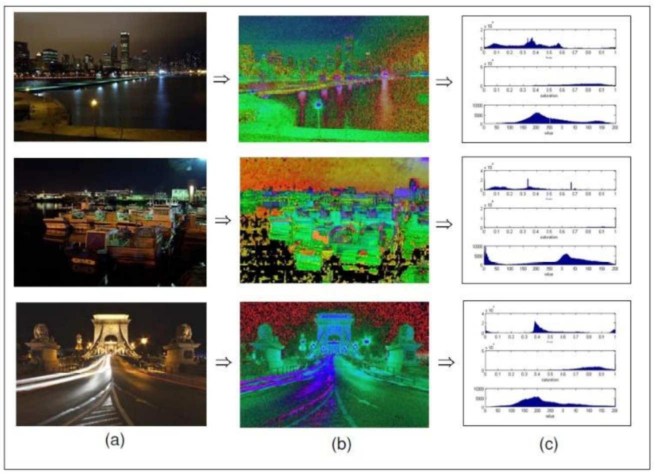

Figure 5. Examples of conversion of color space conversion histogram. Column (a) shows the original images in RGB color space. Column (b) are image in HSV space. Column (c) shows their HSV color histogram

We incorporate knowledge-based expansion into the image retrieval problem because using sub-objects as surrogate terms for general queries is able to improve retrieval accuracy. The presence of particular objects in an image often implies the occurrence of other objects. For example, the concept "wedding" may be expanded to bride, groom, flower and wedding cake. Such knowledge-based expansion allows the semantic content of an image to be automatically expanded.

Our rule-based approach depends on image acquisition parameters to extract scenes of images the photographers intend to capture. In the event that photographers make use of improper acquisition parameters, unexpected content of image would result. To eliminate the noise and enrich the rule-based annotation, we augment the rule-based approach in conjunction with image processing and similarity extraction of image content. From the joint application of these techniques, we can formulate meaningful semantic annotations.

Generation of Mpeg-7 Descriptions

In MPEG-7, the Multimedia Description Scheme (MDS), the MPEG specified Descriptors and Descriptor Schemes dealing with generic features and multimedia descriptions are defined. Objects derived from semantic objects and an extendable set of predefined relations between these objects can be used for constructing a semantic graph, describing the multimedia content [24]. MPEG-7 facilitates Text Annotation Datatypes to support high-level concept enrichment. An MPEG-7 Text Annotation Datatype supports annotations in different forms and allows multiple annotations. In this paper, we focus on the Structured Annotation Datatype.

The Structured Annotation Datatype is one that gives a textual description of events, people, animals, objects, places, actions, purposes, times, attributes and behavior. It provides a structured format that is simple yet an expressive and powerful annotation tool. Syntactic relations such as subject, object and verb modifiers between actions and objects can be described.

By using our method, some of these fields, particularly the first four, may be automatically filled in a meaningful manner. In addition, semantic concepts may be enriched and expanded through semantic manipulation, which may work in conjunction with specialization and generalization hierarchies. Thus, searching for an image with a particular object type may be specialized to a narrower type; for example, in searching for people, the portrait category may be used, and in searching for animals, the wildlife category will fit the requirement. The more extensive and complete such hierarchies, the greater the scope for rich semantic manipulation.

Fig. 6 gives some examples of automatically generated MPEG-7 Structured Annotation descriptions. Consider the third photograph, the application of our rules indicates that this is a portrait image category and scenes of an outdoor event. Thus, "People" and "Portraits" are indicated under WhatObject. As this is an outdoor event, "Outdoor event" is indicated under WhatAction. The location is obtained from the GPS coordinates which map to Cotton Tree Drive Marriage Registry in Hong Kong; thus, in the Where field, the Cotton Tree Drive Marriage Registry is indicated which may be used as a search argument, and in traversing the generalization hierarchy, Hong Kong is also indicated. Timestamp information allows the season, time of day, and year to be variously indicated. Further descriptions are possible, depending on how far up or deep down the relevant hierarchy is being traversed, all of which may be used as search arguments.

Besides the descriptions generated from image metadata, some of the semantic description fields can be further filled with other image features. Fig. 7 gives an example of further enrichment of semantic concepts with face recognition.

To validate our model, we test our system using an evaluation images set. We have evaluated the proposed approach

ISSN:2319-7900

on 100,000+ images of Flickr.com with associated 1,000,000 user's tags. We have manually examined 200 images for each scene and category of images and with total results for 2,600 images. In Fig. 8 shows the accuracy about our model compared with the traditional human tagging model. We found that experimental results of traditional human-tagging are reliable with over 72% accuracy across all scenes and categories. Our model delivers good results and sometimes have better accuracy than human tagging models, especially for the nature categories, with wildlife and macro scenes having 9% to 18% advantage.

Figure 6. Automatic generation of semantic MPEG-7 descriptions for image from metadata: (a) original images (b) automated MPEG-7 descriptions

Fig. 9 gives the experimental results of tagging accuracy. Our experiments indicate that while tags by humans deliver excellent performance with 100% accuracy but such tagging relies heavily on human involvement. Besides, we can obtain highly competent results from our model with HSV color space histogram distances fully automatically where accuracies grow to 67.2% (BD), 71.4% (CD) and 77.9% (ED). For the model combined with global color distribution, the accuracy is 85.3%. While the model incorporating edge detection algorithms, the precision rate is 87.1%. From the joint application of these, we can formulate semantic annotations for specific image fully automatically and index images purely by machine without any human involvement.

Figure 7. An example of automatic generation of semantic MPEG-7 descriptions for image from metadata and low level features

Conclusion

The MPEG-7 Structured Annotation Datatype is often regarded as a powerful semantic information bearing scheme for images and multimedia objects. Through the use of the Structured Annotation Datatype, semantic identification and search of images using arguments meaningful to human beings may be used. In the past, Structured Annotation descriptions were generally handcrafted by humans manually. This has become increasingly impractical due to the rapid rate with which images are captured, created, and uploaded.

By the use of image metadata, augmented with image processing and ontological mechanisms, a methodology has been presented which allows the automatic generation of MPEG-7 Structured Annotation descriptions. Admittedly, in its utmost generality, the full Structured Annotation description is comprehensive and challenging even for humans. However, the present automatic approach is able to go a long way towards providing humanly useful and meaningful descriptions by filling out automatically some of the key fields within the Structured Annotation Datatype. In the same way, as other forms of metadata such as timestamps are currently already stored together with the images, such automatically generated standard semantic fields may also be routinely attached to images to take their search to a deeper and richer content-oriented semantic level. Additional refinement is no doubt possible and desirable in the future to further fine tune

ISSN:2319-7900

and enrich such descriptions, and the present method is able to provide an important first step towards this.

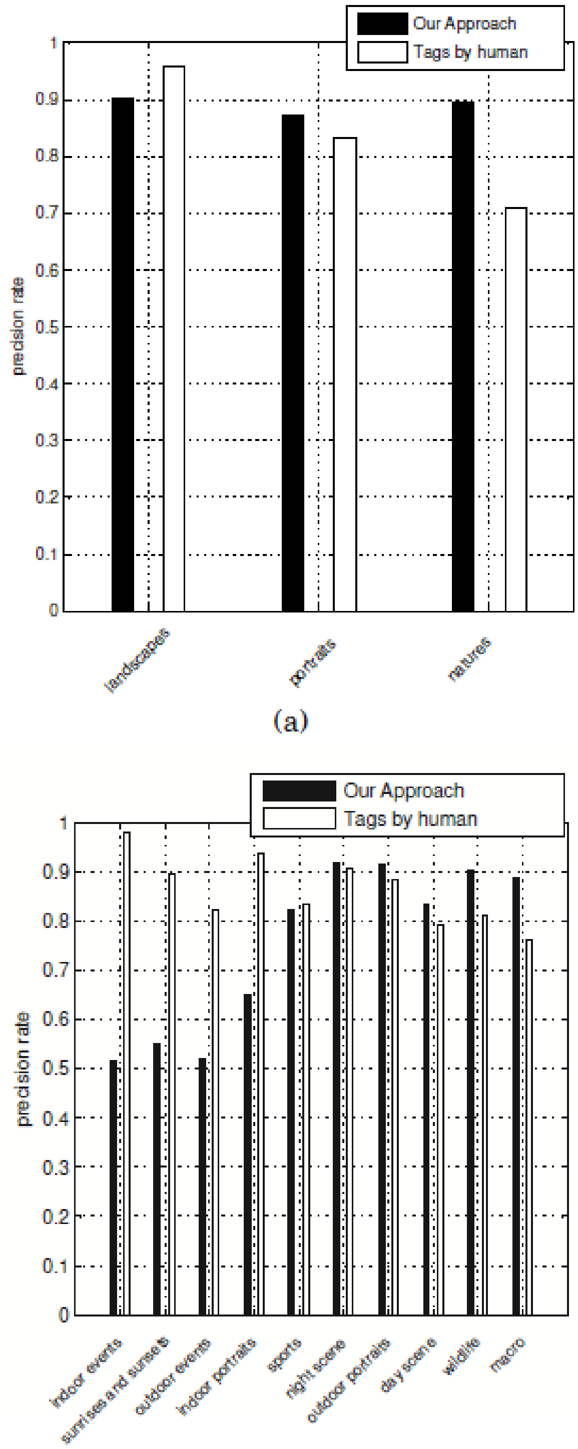

Figure 8. Comparison between the proposed approach, tags by human grouped by (a) scenes of images (b) images categories

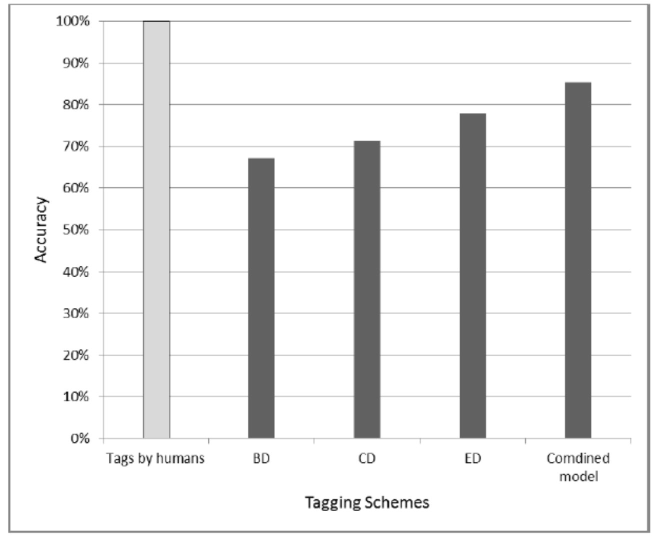

Figure 9. Comparison of experimental results of human tagging and our model with HSV color space histogram distances

References

[1] I. Azzam, A. G. Charlapally, C. H. C. Leung, and J. F. Horwood. 2004. Content based Image Indexing and Retrieval with XML Representations. In *Proceedings of the International Symposium on Intelligent Multimedia, Video and Speech Processing*. Hong Kong, 181–185.

[2] I. A. Azzam, C. H. C. Leung, and J. F. Horwood. 2004. Implicit Concept-based Image Indexing and Retrieval. In *Proceedings of the IEEE International Conference on Multi-media Modeling*. IEEE Computer Society, Los Alamitos, CA, USA, 354–359. https://doi.org/10.1109/MULMM.2004.1265007

[3] I. A. Azzam, C. H. C. Leung, and J. F. Horwood. 2005. A Fuzzy Expert System for Concept-Based Image Indexing and Retrieval. In *Proceedings of IEEE International Conference on Multimedia Modeling Conference*, Vol. 0. IEEE Computer Society, Los Alamitos, CA, USA, 452–457. https://doi.org/10.1109/MMMC.2005.8

[4] K Barnard, P Duygulu, N de Freitas, D Forsyth, D Blei, and MI Jordan. 2003. Matching Words and Pictures. *Journal of Machine Learning Research* 3 (2003), 1107–1135.

[5] Matthew Boutell and Jiebo Luo. 2004. Bayesian Fusion of Camera Metadata Cues in Semantic Scene Classification. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 2. IEEE Computer Society, Los Alamitos, CA, USA, 623–630. https://doi.org/10.1109/CVPR.2004.49

[6] Matthew Boutell and Jiebo Luo. 2004. Photo Classification by Integrating Image Content and Camera Metadata. In *ICPR'04: Proceedings of IEEE International Conference on Pattern Recognition*. IEEE Computer Socie-

ISSN:2319-7900

ty, Washington, DC, USA, 901–904. https://doi.org/10.1109/ICPR.2004.693

[7] June Suh Cho and Joonsoo Choi. 2005. Contour-based Partial Object Recognition using Symmetry in Image Databases. In *SAC'05: Proceedings of the 2005 ACM symposium on Applied computing*. ACM Press, New York, NY, USA, 1190–1194. https://doi.org/10.1145/1066677.1066947

[8] Sonya Coleman, Bryan Scotney, and Dermot Kerr. 2007. Integrated Edge and Corner Detection. In *Proceedings of International Conference on Image Analysis and Processing*, Vol. 0. IEEE Computer Society, Los Alamitos, CA, USA, 653–658. https://doi.org/10.1109/ICIAP.2007.80

[9] Daniel Cremers, Mikaël Rousson, and Rachid Deriche. 2007. A Review of Statistical Approaches to Level Set Segmentation: Integrating Color, Texture, Motion and Shape. *International Journal of Computer Vision* 72, 2 (2007), 195–215.

[10] Ritendra Datta, Weina Ge, Jia Li, and James Z. Wang. 2007. Toward Bridging the Annotation-Retrieval Gap in Image Search. *IEEE MultiMedia* 14, 3 (2007), 24–35. https://doi.org/10.1109/MMUL.2007.67

[11] Ritendra Datta, Jia Li, and James Z. Wang. 2005. Content-based image retrieval: approaches and trends of the new age. In *MIR'05: Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval*. ACM, New York, NY, USA, 253–262. https://doi.org/10.1145/1101826.1101866

[12] Ismail El Sayad, Jean Martinet, Thierry Urruty, and Chabane Djeraba. 2012. Toward a higher-level visual representation for content-based image retrieval. *Multimedia Tools Appl.* 60, 2 (sep 2012), 455–482. https://doi.org/10.1007/s11042-010-0596-x

[13] A. Fakhari and A.-M. Eftekhari-Moghaddam. 2011. Hierarchical Decision Tree (HDT) Approach for Image Annotation. In *Machine Vision and Image Processing (MVIP), 2011 7th Iranian*. 1–5. https://doi.org/10.1109/IranianMVIP.2011.6121600

[14] Lisa Fan and Botang Li. 2006. A Hybrid Model of Image Retrieval Based on Ontology Technology and Probabilistic Ranking. *Web Intelligence* 0 (2006), 477–480. https://doi.org/10.1109/WI.2006.16

[15] Juergen Gausemeier, Juergen Fruend, Carsten Matysczok, Beat Bruederlin, and David Beier. 2003. Development of a Real Time Image Based Object Recognition Method for Mobile AR-Devices. In *Proceedings of International conference on Computer graphics, virtual Reality, visualisation and interaction in Africa*. ACM Press, New York, NY, USA, 133–139. https://doi.org/10.1145/602330.602355

[16] N. V. Hoàng, V. Gouet-Brunet, M. Rukoz, and M. Manouvrier. 2010. Embedding spatial information into image content description for scene retrieval. *Pattern Recogn.* 43, 9 (sep 2010), 3013–3024. https://doi.org/10.1016/j.patcog.2010.03.024

[17] Amit Kale, Kenneth Kwan, and Christopher Jaynes. 2004. Epipolar Constrained User Pushbutton Selection in Projected Interfaces. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 10. IEEE Computer Society, Los Alamitos, CA, USA, 156–164. https://doi.org/10.1109/CVPR.2004.332

[18] William Ku, Mohan S. Kankanhalli, and Joo Hwee Lim. 2007. Using Camera Settings Templates ("Scene Modes") for Image Scene Classification of Photographs Taken on Manual/Expert Settings. In *Proceedings of Pacific-Rim Conference on Multimedia*. 10–17.

[19] Clement H. C. Leung and Roger C. F. Wong. 2009. *Machine Learning Techniques for Adaptive Multimedia Retrieval: Technologies Applications and Perspectives*. IGI Global, Chapter Automatic Image Annotation and Retrieval: Semantic Expansion Approaches.

[20] Jia Li and James Z. Wang. 2008. Real-Time Computerized Annotation of Pictures. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30, 6 (2008), 985–1002. https://doi.org/10.1109/TPAMI.2007.70847

[21] Xiaoyan Li, Lidan Shou, Gang Chen, Tianlei Hu, and Jinxiang Dong. 2008. Modeling Image Data for Effective Indexing and Retrieval in Large General Image Databases. *IEEE Transactions on Knowledge and Data Engineering* 20, 11 (2008), 1566–1580. https://doi.org/10.1109/TKDE.2008.56

[22] H. Lieberman and H. Liu. 2002. Adaptive Linking between Text and Photos Using Common Sense Reasoning. *Adaptive Hypermedia and Adaptive Web-Based Systems, Second International Conference, AH 2002, Malaga, Spain* (May 2002), 2–11.

[23] David Liu and Tsuhan Chen. 2006. Content-Free Image Retrieval using Bayesian Product Rule. In *Proceedings of IEEE International Conference on Multimedia and Expo*, Vol. 0. IEEE Computer Society, Los Alamitos, CA, USA, 89–92. https://doi.org/10.1109/ICME.2006.262557

[24] M. Lux. 2009. An Evaluation of Metrics for Retrieval of MPEG-7 Semantic Descriptions. In *Multimedia, 2009. ISM '09. 11th IEEE International Symposium on* 546–551. https://doi.org/10.1109/ISM.2009.104

[25] J. Manikandan, B. Venkataramani, and M. Jayachandran. 2007. Evaluation of Edge Detection Techniques towards Implementation of Automatic Target Recognition. In *Proceedings of International Conference on Computational Intelligence and Multimedia Applications*, Vol. 2. IEEE Computer Society, Los Alamitos, CA, USA, 441–445. https://doi.org/10.1109/ICCIMA.2007.84

[26] Tao Meng and Mei-Ling Shyu. 2012. Leveraging Concept Association Network for Multimedia Rare Concept

ISSN:2319-7900

Mining and Retrieval. In *Multimedia and Expo (ICME), 2012 IEEE International Conference on*. 860–865. https://doi.org/10.1109/ICME.2012.134

[27] Alexandros Ntousias, Nektarios Gioldasis, Chrisa Tsinaraki, and Stavros Christodoulakis. 2008. Rich metadata and context capturing through CIDOC/CRM and MPEG-7 interoperability. In *Proceedings of the 2008 international conference on Content-based image and video retrieval (CIVR '08)*. ACM, New York, NY, USA, 151–160. https://doi.org/10.1145/1386352.1386377

[28] Richard Pawlicki, István Kókai, Jörg Finger, Randall Smith, and Thomas Vetter. 2007. Navigating in a Shape Space of Registered Models. *IEEE Transactions on Visualization and Computer Graphics* 13, 6 (2007), 1552–1559.

[29] Adrian Popescu, Gregory Grefenstette, and Pierre Alain Moellic. 2006. Using Semantic Commonsense Resources in Image Retrieval. In *International Workshop on Semantic Media Adaptation and Personalization*. IEEE Computer Society, Washington, DC, USA, 31–36. https://doi.org/10.1109/SMAP.2006.37

[30] Muhammad Riaz, Gwangwon Kang, Youngbae Kim, Sungbum Pan, and Jongan Park. 2008. Efficient Image Retrieval Using Adaptive Segmentation of HSV Color Space. In *Proceedings of IEEE International Conference on Computational Sciences and Its Applications*, Vol. 0. IEEE Computer Society, Los Alamitos, CA, USA, 491–496. https://doi.org/10.1109/ICCSA.2008.55

[31] Pinaki Sinha and Ramesh Jain. 2008. Classification and annotation of digital photos using optical context data. In *Proceedings of the International Conference on Image and Video Retrieval*. ACM, New York, NY, USA, 309–318.

[32] G. Stamou, J. van Ossenbruggen, J.Z. Pan, G. Schreiber, and J.R. Smith. 2006. Multimedia annotations on the semantic Web. *MultiMedia*, *IEEE* 13, 1 (2006), 86–90. https://doi.org/10.1109/MMUL.2006.15

[33] A. M. Tam and C. H. C. Leung. 2000. Semantic Content Retrieval and Structured Annotation: Beyond Keywords. *ISO/IEC JTC1/SC29/WG11 MPEG00/M5738, Noordwijkerhout, Netherlands* (March 2000).

[34] A. M. Tam and C. H. C. Leung. 2001. Structured natural-language descriptions for semantic content retrieval of visual materials. *Journal of American Society for information Science and Technology* 52, 11 (March 2001), 930–937. https://doi.org/10.1002/asi.1151.abs

[35] V.S. Tseng, Ja-Hwung Su, Hao-Hua Ku, and Bo-Wen Wang. 2008. Intelligent Concept-Oriented and Content-Based Image Retrieval by using data mining and query decomposition techniques. In *Multimedia and Expo, 2008 IEEE International Conference on*. 1273–1276. https://doi.org/10.1109/ICME.2008.4607674

[36] P. van Beek et. al. 2001. Multimedia Content Description Interface - Part 5 Multimedia Description Schemes. *ISO/IEC JTC1/SC29/WG11 MPEG01/N3966, Singapore* (March 2001).

[37] Julia Vogel, Adrian Schwaninger, Christian Wallraven, and Heinrich H. Bülthoff . 2007. Categorization of natural scenes: Local versus global information and the role of color. *ACM Transactions on Applied Perception* 4, 3 (2007), 19. https: //doi.org/10.1145/1278387.1278393

[38] Jun Wang, Jia Bei, Tongwei Ren, Yan Liu, and Bowei Xu. 2012. Image retargeting assessment based on salient region similarity. In *Proceedings of the 4th International Conference on Internet Multimedia Computing and Service (ICIMCS '12)*. ACM, New York, NY, USA, 140–143. https://doi.org/10.1145/2382336.2382376

[39] Xin-Jing Wang, Lei Zhang, Xirong Li, and Wei-Ying Ma. 2008. Annotating Images by Mining Image Search Results. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30, 11 (2008), 1919–1932. https://doi.org/10.1109/TPAMI.2008.127

[40] Roger C. F. Wong and Clement H. C. Leung. 2008. Automatic Semantic Annotation of Real-World Web Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30, 11 (November 2008), 1933–1944. https://doi.org/10.1109/TPAMI.2008.125

[41] Thomas Zöller and Joachim M. Buhmann. 2007. Robust Image Segmentation Using Resampling and Shape Constraints. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29, 7 (2007), 1147–1164. https://doi.org/10.1109/TPAMI.2007.1150

Biographies

Prof. Clement H.C. Leung is a Professor in Computer Science and Technology, United International College. He obtained his BSc (Hons) in Mathematics from McGill University, Canada, his MSc in Mathematics from Oxford University, and his PhD in Computer Science from University College London. His principal research interest is in the area of Multimedia and Visual Information Systems. He holds two US patents, and his publications include four books and well over one hundred research articles. Among the publications that his articles have appeared include: IEEE Transactions on Pattern Analysis and Machine Intelligence, ACM Journal of Multimedia, IEEE Transactions on Communications, ACM Transactions on Intelligent Systems Technology, ACM Computer Graphics, The Computer Journal, IEEE Transactions on Software Engineering, and IEEE Multimedia. He is acknowledged by the international research community to be a pioneer in the field of Visual Information Systems, and has published the first research volume on the subject. His services to the research community include serving as Program Chair, Program Co-Chair, Keynote Speaker, Panel Expert, and on the Program Committee and

ISSN:2319-7900

Steering Committee of major International Conferences. In addition to contributing to the Editorship of ten international journals, he has served as Chairman of the International Association for Pattern Recognition Technical Committee on Multimedia and Visual Information Systems, as well as well as on the International Standards (ISO) MPEG-7 Committee responsible for generating standards for digital multimedia, where he played an active role in shaping the influential MPEG-7 International Standard. He is listed in Who's Who in the World, Great Minds of the 21st Century, Dictionary of International Biography, Who's Who in Australia, and Who's Who in Australasia & Pacific Nations. He is a Fellow of the British Computer Society, and a Fellow of the Royal Society of Arts, Manufactures and Commerce.

Dr. Yuanxi Li is currently a lecturer at Department of Computer Science, Hong Kong Baptist University. She received her Ph.D. degree in Computer Science and M.Sc. in IT Management from Hong Kong Baptist University. Her research interests include Semantic Multimedia Indexing and Retrieval, Adaptive Searching Engine and Semantic Ontology, etc. She has been serving as committee in several International Conferences and Workshops, including organizing committee RTCSA2015, co-chair of UCC-IDP2016, etc. Besides, she has been also reviewers for several Journal and conference, including the International Journal of Web Information Systems, the International Conference on Utility and Cloud Computing, etc.