Image Concepts Disambiguation using Associated Text Concepts

Ahmad Adel Abu Shareha, Rajeswari Mandava, Dhanesh Ramachandram
Computer Vision Research Group, School of Computer Science, Universiti Sains Malaysia
11800, Penang, Malaysia
{adel,mandava,dhaneshr}@cs.usm.my

ABSTRACT

A technique for image concepts disambiguation using associated textual information is presented as an extension to our previous work on image-text integration [1]. The integration results are used to set the image and text into common ground. Then, the processed textual information is used to disambiguate the image concepts using confidence updating in Bayesian Network. Since mostly, image and text have no direct corresponding, the extracted concepts from image and text are scrambled out in Bayesian Network to show the influence factor of these concepts on each other. The prior probability of the BN node has been calculated to address the bilateral direct and indirect relationships between the concepts while, CPTs are built to address the mixture relationships of the entire modals (image and text). Thus, the final results represent the effects of all the concepts of both sides on the image concepts.

Key Words: Semantic image analysis, Image concept disambiguation, Image-text integration

1. Introduction

In recent developments in the image processing field, extraction semantic content information about the image has become a requirement in various applications, especially in image indexing and retrieval. [2, 3] Thus, the extracted machine low-level features have to be linked to the high level semantics concepts; this is what so called the image concept extraction. In general, concept extraction is a challenging task which is basically achieved by mapping the machine extracted low-level features into their corresponding stored labels (feature-to-label).

Image disambiguation is responsible for disambiguation of the ambiguous concepts produced by the concept extraction. Using some information about the image and the produced concepts, the disambiguation process refines such ambiguity and identifies the projected concept. There are several particulars that can be used in this sense such as content information, context information, geometric information and any available augmented data. Disambiguation using image-associated information is a promising approach because; most probably the ambiguity problem of the image information will not occur in the augmented information. In this sense, the most generic informative sources associated with the image in the real application are the text. In the literature, the textual information such as: captions, annotation and keywords have been utilized for the disambiguation purpose [4]. However, free text which is more likely to be associated with the image has not been utilized yet. This paper presents a method for image concepts disambiguation using associated free text information as an extension to our previous work on image-text integration. The integration results are used to set the image and text into common ground. Then, the processed textual information is used to disambiguate the image concepts.

The logical combination of image-text association has a lot of possibilities (ex: projection or expansion which in turn can be an elaboration extension or enhancement extension).[1] Therefore, an image and its

associated text are expected to have different generality, abstraction and coverage levels although they express the same meaning. These differences have been addressed in the image-text integration. Thus, an image disambiguation using associated text is possible relying on such integration. In our previous work [1], the image and the associated text are considered as two modals which processed independently in the aid of the knowledge that is encoded in ontologies, these modals are represented as two list of concepts. The integration process commences by utilizing the ontologies to enrich both lists which resolve the coverage differences problem and maximizes their incorporation. Then, the enriched lists are integrated. This integration relies on the ontology alignment which matches up the underlying ontology. The ontology alignment solves the granularity level differences problem and sets better matching than this can be achieved by matching the lists directly, that is because the ontologies have richer structure. However, the integration of image and text are based on the ambiguous image concepts. Thus, this paper utilized the integration results to disambiguate the ambiguous image concepts using confidence updating in Bayesian Network.

2. Previous Work

Mostly, the disambiguation techniques are presented as a part of concept extraction systems. There are different types of particular information that can be used for the disambiguation purpose such as content information, context information, geometric information and any available augmented data. In the context-based approach, information about the scene of the image is used to disambiguate its content, under the assumption that 'more likely' or 'less likely' that one object appears in specific context. Ph. Mylonas et al. [5, 6] have utilized the contextual information that is obtained from domain-specific ontology to re-adjust the labeling information that is extracted using content analysis and the concept extraction system. Each of the extracted concepts has context relevance value. The goal is to select a concept or group of concepts with the higher context relevance.

In the augment data based approach, Benitez, A.B. and S.-F. Chang [4] used WordNet to disambiguate annotated images by linking the extracted concepts from the underlying image with some keywords that was extracted from its annotation text. Each concept from the image is disambiguated by seeking out the annotation text for a corresponding word throughout WordNet. Matching keywords directly can be sufficient in the case that the annotation text is carefully describing the content information of the image. This, in turn, needs to restrict the annotator with tedious rules which are mostly impractical. On the other hand, this paper introduced a technique for disambiguating concepts which does not match keywords directly. The presented technique, can process any form of image-text association without restriction on the text presentation.

Image-text association is common in many real-world applications, such as scientific documents, web pages and newspaper advertisements. The processing of such association has been investigated from different prospective such as classification, clustering, indexing and retrieval. In the classification problem for example, a high dimensional space is constructed normally to represent the textual feature (words) and the visual feature (low-level) [7]. Using another technique, two different vectors are constructed separately for each data source. Then, the similarity of each vector and the query association is calculated separately and combined linearly. [8, 9] Utilizing the low level features as mentioned earlier does not satisfy the demands of having semantic output. Thus, semantic information extraction from an image-text association can be utilized in the previously mentioned real-world application. This paper presented

a disambiguation of image concepts using free textual information based on high level image-text integration. The disambiguation process is considered as an important step towards semantic extraction which can be used with different image-text applications and for different purposes.

3. Proposed Work

The proposed technique for image disambiguation relies on the concept extraction process, the utilized ontologies, image-text integration and Bayesian Network.

3.1. Concept Extraction and Ontologies

As a matter of fact, the disambiguation process is an extension to the concept extraction method. In order to develop a good disambiguation technique, it is necessary to understand the scenarios that occur in the extraction processes. Besides, it is necessary to determine and consider the possible forms of the extracted concepts which are going to be the inputs to the disambiguation technique.

Image concepts extraction process maps the low-level features that are extracted from the segmented object into concepts. In reality, multiple objects might share the same low-level features. Thus, multiple mapping can be produced for a single object. [10] In the second scenario which occur in multi-objects images, if no accurate segmentation is available for the underlying domain, which is mostly true, then multiple segmentation possibilities might be produced, for each segmentation possibilities there is multiple mapping which creates more complex output. For both scenarios, the real identity of the object being identified/recognized might not be retrieved due to some physical circumstances like light and camera angel.

However, to facilitate a good concept disambiguation mechanism, the problem of disambiguation has to be clarified. Thus, although it's not always true, we assume that in specific domain, concept extraction can retrieve a list of concepts which include the true objects identity. Following the stated assumption, the disambiguation method aims at elimination of the ambiguous concepts into the underlying true object identity.

The concept extraction requires knowledge source which links the low level features to high level semantics in the image side. The concepts extraction from text required a control vocabulary. Constructing a knowledge base in some domains is an enormous and tedious task. Thus, utilization of existing knowledge sources if available is a better alternative. In our previous work [1] we utilize the existing knowledge that is encoded in ontologies. A well-known and common example of such knowledge source is WordNet [11] and OntoWordNet [12]. Existing image knowledge bases and sources are less, simply because the image knowledge is more complicated and nonstandard. However, the emergence of the ontology manufactured has supported the representation of such complex image knowledge [5, 6], that is because ontology has well defined structure, strong semantic and reasoning support. FMA which encodes the anatomical structure in the medical domain can be used in our case. [13]

3.2. Image-Text Integration

The image- text integration process takes the inputs as list of concepts from the image and list of concepts from text. The output is correlation information which shows the points of similarities and differences between the two modalities despite their coverage scope, granularity details and generalization level. Figure 1, illustrates the integration process which involves the following steps:

1. Concept mapping and enrichment: This step aims at mapping the concepts into their corresponding ontology entities and enriching them with higher level concepts. The mapping and enrichment shrinks the difference between the image and the associated text in term of coverage.

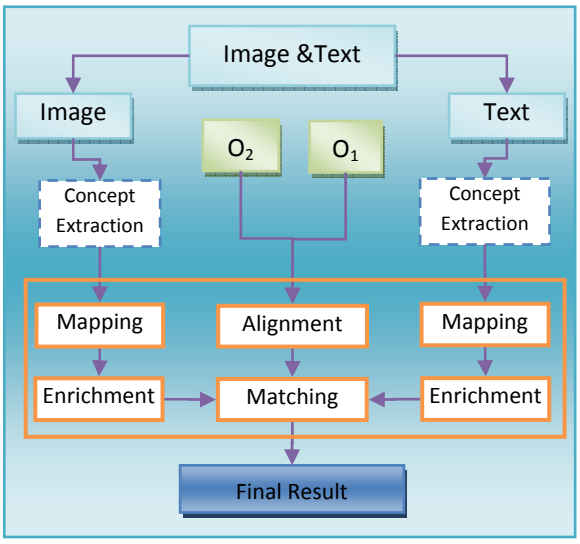

Figure 1: Image-Text Integration

For example, the image concept might be "human" while the text concepts might be "boy". 'human' and 'boy' are referring to the same thing with different abstraction levels. Figure 2, shows an example of mapping and enrichment process.

2. Ontology Alignment: This step aims at matching the entities of the underlying ontologies and discovers their identical elements. To overcome the representation differences between the modalities. The alignment produces an intermediate dialogue which maximizes the ontologies integration.

3. Cluster Matching. This step aims at extracting the output results by showing the scope and the individual similarities between the image and the text modalities. [1]

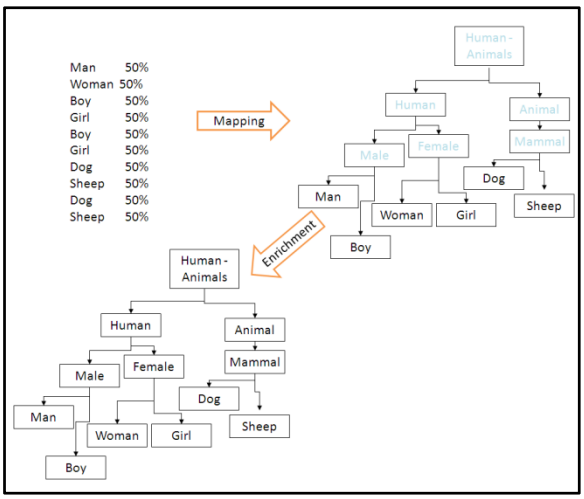

Figure 2: Mapping and Enrichment

Thus, such output can be utilized for any further process such as disambiguation which is addressed in this paper.

3.3. Image Concepts Disambiguation

As mentioned earlier, the image concepts are produced by mapping the low-level features to its identity. In the simple case of single-object images or for multi-objects images wherein accurate segmentation is available, the mapping produces multiple labels for each single object each with a confidence value. In such cases, the disambiguation is achieved by directly matching any of the ambiguous concepts with some similar text concept, or by looking for the shortest distance between the text concepts and any of the ambiguous image concepts in the underlying ontologies.

Intuitively, the image and the text representation have no direct correspondence. The reasons of such indirect correspondence are the factors of the differences that are mentioned earlier which are representation, granularity and abstraction differences besides, two or more concepts from the text might be correlated to support a single image concept and finally, the poor segmentation methods available for the underlying domain. Thus, in such case there is a need to process all the concepts from both modalities together. To address such complex condition, the concepts are scrambled out in structure representation to show the influence factor of these concepts on each other. The proposed technique uses Bayesian Network for this purpose. The constructed BN is able to process both direct and indirect image-text correspondence. The prior probability of the BN node has been calculated to address the bilateral direct and indirect relationships between the concepts while, CPTs are built to address the mixture relationships of the entire modals (image and text).

3.31. Bayesian Network

Bayesian belief network is a graphical representation which constructs conditional,

influences and probabilistic relationships between variables. Bayesian Network can handle inexact information efficiently better than any other artificial intelligent formalism. Thus, it is used with the proposed disambiguation process. Bayesian network is created for the ambiguous data with structure equivalent to the ontology structure in the image. The nodes in BN are corresponding to the ontologies concepts in the image side. The relationships between these concepts follow the relationships in the original ontology. The text- side extracted concepts that have corresponding ontology entities which are aligned to the image-side ontology entities are also mapped to nodes in BN. Thus, the concepts that have no alignment dialogue should be skipped because it might cause an inconsistency problem in the underlying structure when the relationships are built. The edges in BN are corresponding to the concepts relationships at image side to avoid conflicting between image and text. After performing these steps, the BN structure is ready to be used. The nodes of the BN are now representing the scrambling of the image and text concepts in unified representation because of the alignment process. Then, the prior probability of each node is calculated using the following formula.

$$P(N_i^B) = (\alpha\, P(C_i^I) + \beta\, P(C^T_{i\,successor}) + \delta) \in (0,1]$$

where, $P(N_i^B)$ is the prior probability of the node i in BN. $P(C_i^I) \in [0,1]$ is the confidence value of the concept i in the image side. $P(C^T_{i\,successor}) \in \{0,1\}$ is the confidence value of the concept i or its successor in text side. Because, recognize the child concept is an indicator of having the ancestor. For example, 'Dog' is an indicator of 'Animal'.

$$\alpha, \beta, and\ \delta\ are\ constants\ \delta > 0\ to\ avoid\ zero\ values$$

After building the BN, creating the nodes and calculating the prior probability based on the provided formula, each of the nodes will have the exact impact from the image side and the related impact from the text side. Thus, each node has a confidence value represents a combination of image-text confidences.

3.3.2. Conditional Probability Table CPT

The CPT represents the way in which the confidence value of a node affects the confidence value of the connected nodes. CPTs are a set of Meta-rules which depend on the experience or history of the problem and/or the domain. In the disambiguation process, CPT represents the mixture effects of the concepts at different levels of abstraction on each other. The rules that used here are represented as:

$$N_i^B ischildof\ N_j^B \rightarrow P(N_i^B|N_j^B) = \left(\left(\frac{P(C_i^I)}{No\ of\ Siblingsof\ C_j^I}\right) + P(C_i^T)\right) \in (0,1]$$

Accordingly, the confidence value of the image concepts is distributed into its child nodes equally. If an 'Animal' where extracted from image then the chance for this animal to be 'Tiger' or 'Dog' is equal, because both of them are considered as an animal. On the other hand, it gives direct impact to the text concepts.

As the BN run the influence of the text concepts which is not related directly on the image concepts and the influence of the image concepts on each other is taken place. Thus, the final results represent the effects of all the concepts in both sides on the image concepts. Finally Bayesian Network run to update the confidence values of the image concepts. The nodes with higher confidence value are considered to be the true objects.

4. Conclusion

This paper proposed an image concepts disambiguation technique which utilizes associated text concepts. The input concepts from both image and text are injected into a Bayesian Network. After building the BN, creating the nodes and calculating the prior probability based on the provided formula,

each of the nodes will have the exact impact from the image side and the related impact from the text side. Thus, each node has a confidence value represents a combination of image-text confidences. As the BN run the influence of the text concepts, which are not related directly to the image concepts and the influence of the image concepts on each other is taken place. Thus, the final results represent the effects of all the concepts in both sides on the image concepts. Finally Bayesian Network updates the confidence values of the image concepts. The nodes with higher confidence value are considered to be the true objects.

References:

[1] A.A. Abu Shareha, M. Rajeswari, and D. Ramachandram, "Multimodal (Image & Text) Integration Using Ontology Alignment," American Journal of Applied Science, vol. 6, no. 6, 2009, pp. 1217-1224.

[2] R.S. John, and C. Shih-Fu, "VisualSEEk: a fully automated content-based image query system," Proceedings of the fourth ACM international conference on Multimedia, ACM. Boston, Massachusetts, United States, 1992, pp. 87-98

[3] P.P. Alexander, W.P. Rosalind, and S. Stan, "Photobook: tools for content-based manipulation of image databases," SPIE, 1995 pp. 37-50.

[4] A.B. Benitez, and S.-F. Chang, "Semantic Knowledge Construction From Annotated Image Collections," ICME, Lausanne, Switzerland, 2002.

[5] P. Mylonas., T. Athanasiadis, and Y. Avrithis, "Improving Image Analysis Using a Contextual Approach. 7th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2006), Seoul, Korea, 19-21 April, 2006.

[6] P. Mylonas, Th. Athanasiadis and Y. Avrithis, "Image Analysis Using Domain Knowledge and Visual Context", 13th International Conference on Systems, Signals and Image Processing (IWSSIP 2006), Budapest, Hungary, 21-23 September 2006

[7] W.I.G. Rong Zhao, "Narrowing the semantic gap improved text based web document retrieval using visual feature," IEEE Trans. on Multimedia, 4(2):189-200, 2002.

[8] Z. Chen, L. Wenyin, F. Zhang, M. Li, and H. Zhang, "Web Mining for Web Image Retrieval," J. Am. Soc. Information Science and Technology, vol. 52, 2001, pp. 831-839.

[9] T. Gevers, R.F. Aldershoff, and A.W.M. Smeulders, "Classification of images on the internet by visual and textual information," SPIE, Bellingham: Society of the Photo-Optical Instrumentation Engineers, San Jose pp. 16-28.

[10] K.-W. Park, and D.-H. Lee, "Full-Automatic High-Level Concept Extraction from Images Using Ontologies and Semantic Inference Rules," ASWC, pp. 307-321.

[11] G.A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. Miller, "Introduction to WordNet: An On-line Lexical Database," International Journal of Lexicography, vol. 3, no. 4, 1990, pp. 235-244.

[12] A. Gangemi, R. Navigli, and P. Velardi, "The OntoWordNet Project: Extension and Axiomatization of Conceptual Relations in WordNet," On the Move to Meaningful Internet Systems (OTM2003), Springer-Verlag, pp. 820-838.

[13] C. Rosse, and J.L.V. Mejino, "A Reference Ontology for Bioinformatics: The Foundational Model of Anatomy," Journal of Biomedical Informatics, vol. 36, 2003, pp. 478-500.