An Approach to Information Retrieval from Construction Images

Aqli Mursadin
Engineering Faculty, Lambung Mangkurat University
Engineering Faculty Building, Unlam Campus, Banjarmasin 70123, Indonesia
a.mursadin@teknik-unlam.ac.id

Abstract: Construction images are potentially valuable sources of information. However, there are several problems related to its retrieval mainly relating to understanding of the semantics of the corresponding content. Additionally, the information is very much unstructured in nature. These problems have been solved by the development of concepts enabling the content to be represented in structured form. The framework proposed for this is described. It is based on the use of domain-specific concepts to retrieve image content and contextual information to improve content understanding. A mathematical framework is also developed based on a Bayesian approach to make the implementation possible for construction image classification. This robust framework enables fusion of related domain knowledge with new information in a versatile manner. An example demonstrates how it is possible to combine information from both available image data and domain knowledge to understand construction image content.

Keywords: construction image, contextual information, domain knowledge, domain-specific concept, image content, information retrieval, information technology

1. Introduction

Photographic images have been used as information sources in construction projects for various purposes such as project documentation, progress monitoring and work inspection. Construction images contain a wealth of information about construction practices. In particular, they provide information concerning the construction method, the intended construction product, the materials and other resources, and the condition of the site.

Apparently, as sources of information, construction images could be valuable in decision making, especially during the early stages of construction projects, where information is often scarce. Simoff and Maher [1] considered the importance of multimedia as information sources during design processes. With the increasing reliance on information technology in construction and the availability of relatively affordable digital imaging devices, have come possibilities as well as challenges to effectively and efficiently make use of such valuable sources of information.

In principle, a source of information can be useful, if there is a practical way to retrieve meaningful information from it. Images contain information in an unstructured form. This unstructured nature results in difficulties of retrieving meaningful information based on the image content.

In digital images information is stored as pixels. This structure of information does not sufficiently represent its meanings and therefore does not lead to understanding of its content. As a result, despite some advantages from handling digital images as computer files, there is the so-called *semantic gap* between human visual perception and pixel-based information processing. It is in contrast to structured information stored in relational database tables, where the information structure explicitly represents its meanings.

In the area of image processing and computer vision, advances have been made in image understanding, such as the increasing capabilities in object recognition and image classification for natural scene. An essential mechanism in image understanding is the incorporation of *domain-specific concepts*. As concepts convey human understanding about the corresponding domains, domain knowledge becomes important. The roles of domain knowledge can be seen not only from the introduction of concepts to represent image content, but also in its ability to examine content in a particular context [2].

This paper proposes an approach to information retrieval from construction images. This approach is based on the integration of domain-specific concepts (conceptual information) and contextual information to represent construction image content. It emphasises the role of construction domain knowledge in understanding content. The retrieval approach can be seen in the context of construction image understanding, one application of which can be found in image classification. An example is presented to demonstrate how the framework can be implemented. Some further direction is given.

2. Conceptualisation of Image Content

Conceptualisation is a fundamental step in introducing structure to image content. Soo *et al.* [3] demonstrated the use of domain-specific concepts for information retrieval from images to introduce concepts for representing content. Fan *et al.* [4] implemented simple concepts for detecting salient objects in natural images. Luo *et al* . [5] also introduced techniques for retrieving and understanding unconstrained photos. Once again, they showed the need for representing content with concepts.

In construction efforts have been made to introduce structured representation of information. Tatum [6] proposed an information classification system for construction technology. Hanlon and Sanvido [7] developed a classification system for constructability information. A number of construction product information classification systems are also available such as Master Format and Uniclass. Their conceptualisations are not based on visual perception. Also, although some concepts in these classification systems might be adaptable in describing image content, their hierarchy is far too complicated for relatively straightforward information content.

Image understanding requires knowledge about the relationships between concepts leading to the need for ontology. Domain knowledge is particularly important in developing ontology [8]. Simoff and Maher [1] discussed ontology in multimedia data mining for retrieving design information. Their ontology is based on predicate structures to capture information content. Although they only explored data mining from textual information sources, their model may also be useful for information retrieval from construction images.

This paper is part of an ongoing research within which a conceptualisation framework is being developed in relation to image retrieval. It applies some high-level concepts to represent objects whose visual features form recognisable patterns. For instance, "column" and "brick" can be considered as concepts. These objects have a number of visual features recognisable in construction images. Once objects are recognised, they should be linked to each other to establish some known and meaningful relationships. Such relationships should be predefined, and domain knowledge in construction is required for that purpose. Although it is possible to bring this conceptualisation into some explicit ontology, it is not the intention of this paper to present an exhaustive conceptualisation framework.

3. The Proposed Approach

The proposed approach is based on the use of domain-specific concepts to retrieve image content and contextual information to improve content understanding. The framework for construction image understanding is shown in Figure 1. A mathematical framework is also developed to make the implementation possible for image classification. This robust framework enables fusion of related domain knowledge as well as new information in a versatile manner.

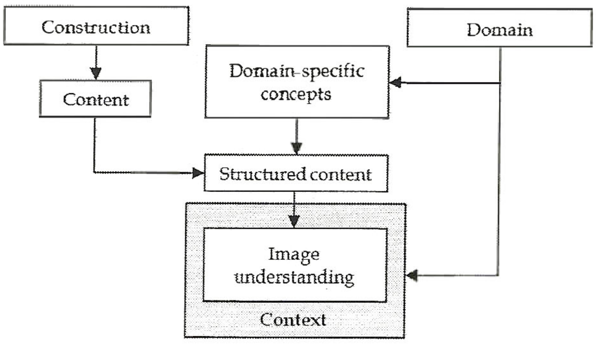


Figure 1. A Framework for Understanding Construction Image Content

3.1 Using Concepts to Capture Content into A Structured Form

Construction images can be seen as scene representations from construction sites. They usually contain complex arrangement of objects. However, unlike natural scene images, construction images are not completely unconstrained and most of the objects in the scenes are consistent with the themes. On the other hand, in understanding construction images one may exclude some objects without significantly affecting the reasoning.

A small pilot survey was conducted involving a few experts knowledgeable in construction. These experts were given a number of construction images and asked a number of questions. For example, they were asked what construction activity is depicted in the image in Figure 2.

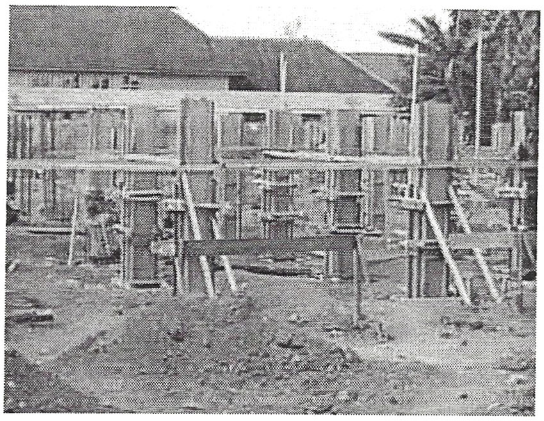


Figure 2. An Example Image

They showed agreement that this is more or less "form installation/setting for columns". Basically, the theme should be related to formwork used for columns. As part of their reasoning they noticed the presence of steel rebars vertically installed and explained that the vertical forms were very likely to imply reinforced concrete columns. They also recognised that the forms were made of timber. On the other hand, the layout of the columns was, according to them, very likely to imply a building. Some even further suggested that the relatively small sizes of the columns and the use of timber forms should imply a not very heavily reinforced concrete structure, which in turn led to a relatively small low-rise building.

From this example one should see a number of important concepts such as "building", "column", "form", "rebar", "concrete", "steel" and "timber". The relationships among these concepts are quite obvious such as "form made of timber", "form used for column", and "column part of building". Interestingly, the presence of some human-like objects was not so significant to them. Some were not even sure what that was and simply ignored its presence in deriving their inference.

There are 3 important aspects in scene (e.g. construction site scene) perception from images. Firstly, some basic concepts/objects should be identified along with the relationships among them. This is important for bridging the semantic gap between low-level vision and high-level perception. Secondly, complete identification of all objects in the scene is not necessary to obtain a reasonably accurate representation. A schematic representation seems to have been more important than detailed recognition of objects in scene perception [9]. Thirdly, domain knowledge is important not only for introducing concepts to capture the content but also for building a schematic representation.

A more structured form can now be introduced to capture high-level information content of images. In this paper a triple is used to represent one *content item*, namely <predicate, concept1, concept2>. This triple consists of the concepts "concept1" and "concept2" and the "predicate" that relates the concepts to each other. For example, the triple <made_of, form, timber> means "form made of timber". Such structures are commonly found in many works related to knowledge representation. Simoff and Maher [1] proposed a similar form for retrieving design information from multimedia sources. A complete conceptualisation framework, which would include an exhaustive library of concepts and their predicates, is being developed and beyond the scope of this paper. Here, only the basic ideas of this framework are discussed.

The following categories of concepts can be used, namely product (P, such as building, column and rebar), material (M, such as concrete, steel and timber) and resource (R, such as formwork). Predicates such as "made_of", "part_of", and "used_for" are included to represent relationships between concepts. Other predicates are also possible. Some possible forms of content items are <made_of, P, M> or <made_of, R, M>, <part_of, P, P> or <part_of, R, R>, and <used_for, R, P>.

This conceptualisation is rather ideal. In reality, it is very difficult to recognise objects in construction images. Such images often contain a very complex scene. Differences between some objects may be too subtle and some images may contain noise, which leave even human with a great deal of ambiguities about the content.

Object identification processes rely heavily on the selection and extraction of low-level visual features (such as colour, texture, and shape). These operations are both computationally exhaustive and hardly optimal. There is a chance that the corresponding vision system recognises the timber form as "timber column" which in turn leads to classifying the image as depicting "a timber structure" instead of "timber forms for concrete columns".

However, as mentioned earlier, complete identification of all objects in the scene is not necessary to obtain a reasonably accurate theme identification and image classification. The context built among content items and according to which they are related to each other can help determine the likelihood of one or more content items. Context incorporation is essential for ensuring accurate perception [2]. It suggests that some content items bring contextual information that helps reduce ambiguities concerning the corresponding image content.

3.2 Incorporating Contextual Information

Context is essential to bring sense concerning content items. Some content items may lead to ambiguities in deciding what they are. As mentioned earlier, there is a chance that the timber forms in Figure 2 are mistakenly perceived as timber columns. However, the presence of steel rebars could reduce such an ambiguity. Also, the joint appearance of timber forms and steel rebars is more likely to lead to a content that makes sense than is the combination between timber columns and steel rebars.

Contextual information among low-level visual features has been utilised to deal with ambiguities in object recognition with plausible results [2]. Knowledge on the meaning of a scene (which is contextual) is also important in the recognition of the constituent objects [10]. Suppose that a context related to a construction of concrete columns is used in retrieving the content from the image in Figure 2, then the item <made_of, form, timber> is more likely to be retrieved than is the item <made_of, column, timber>.

This paper introduces the term *content class*. A content class is related to a number of content items. Image content, as an instance of a content class, could be anything depicted in construction images, from a simple object to a depiction of construction activity to the theme of an image, as long as there is a representation by the corresponding content items. A context is conveyed among its content items. For example, the image in Figure 2 has content that is represented by, at least, <made_of, form, timber> and <made_of, rebar, steel>. Either one of these items conveys contextual information for the content.

Suppose that one asks whether an image has content from a content class *C* that is related to a set of content items. To answer this, the structure of *C* has to be predefined, and then the content items retrievable from the image are examined to see how much they represent *C*. In other words, to retrieve information from images, one should retrieve it based on a particular context conveyed within a predefined content class. The use of context helps reduce the need for complete recognition and combat ambiguities. While in reality construction images may consist of a large number of content items, the use of context helps one focus on more relevant or consistent content items.

3.3 An Application in Construction Image Classification

One particular problem area on which this approach can be applied is image classification where it is decided whether or not an image can be classified as having some particular content. An image classification capability is essential in image retrieval, which in turn is important to ensure the availability of construction images as information sources. With an increasing number of images in collection the problem is how to retrieve the relevant images.

A number of approaches have been proposed for image retrieval by incorporating image understanding, mostly in areas outside construction. Commercial image search engines in the Internet usually base their search on image file names or the surrounding text. The reliability of such a method is not satisfactory as it is the authors of the files or the descriptions who understand the content, and not the search system itself.

The use of metadata is also limited to information having little to do with the content such as the title, the date/time and the author, while it is the conceptual information that is the most important to human [11]. Including a brief description in metadata is not a good practice for image retrieval as once again it is the author not the retrieval system that understands the content. Consequently, the retrieval results may contain images with poor relevance to the user query.

To improve relevance other approaches are based on relevant feedback. Using such feedbacks from users the retrieval system could refine the search results. Wenyin *et al.* [12] proposed the use of relevant feedback for semi-automatic image annotation. Relating words to image contents are the basic idea behind image annotation. By relevant feedback users are expected to select/deselect words based on their relevance to image contents. The annotation results could therefore be useful in image retrieval. By providing initial images as examples, the blind relevant feedback approach proposed by Brilakis and Soibelman [13] could improve similarity between the results and the examples. User intervention was also claimed as minimum in this approach.

Although there would be an implicit image understanding in relevant feedback, consistency may still become an issue. In practice, one should not expect users to provide a reliable judgement on the relevance of an image in a short period of feedback-giving session. A recent study by Chua *et al.* [14] has found significant evidence that people raised in one particular culture (such as the eastern culture) may perceive images in different ways to people raised in another culture (such as the western culture). Since the feedbacks should also be accumulated leading to improvement in future performance, there should be a learning mechanism performed by the system. However, learning based on feedback is usually unstable. Overall, the above approaches are not based on explicit conceptualisation in understanding image content.

The proposed framework, on the other hand, enables the degree that an image has some particular content to be determined based on the extent to which its content items represent the target content; and this means a transparent and traceable mechanism. An application in construction image classification is formulated below.

Let a content class C be represented by a set of content items, $\{c_i\}$, where c_i is the i^{th} content item, $i = 1, \ldots, n$, and n is the number of content items. The extent to which an image can be said as having some content from a class C is commensurate with the extent to which its content items represent C. This requires that C is predefined.

Consider X_i as a variable whose value represents the presence or absence of c_i in an image, then X_i is a random (some may prefer the term 'uncertain') variable. For the sake of formality, letting a real number x_i be 1 or 0 enables one to assign this number to represent X_i, for which $x_i = 1$ means c_i is present and $x_i = 0$ means c_i is absent. Further, it is assumed that these values are available as the results of object recognition.

From elementary statistics, the extent to which an image has some content of class C, given that some of its content items are represented by $X_1, \ldots, X_n$ is the conditional probability value

$$P(X_{n+1} | X_1, \ldots, X_n) = \frac{P(X_1, \ldots, X_n, X_{n+1})}{P(X_1, \ldots, X_n)} \quad (1)$$

where X_{n+1} represents whether the corresponding image contains C (= 1) or not (= 0). This basic formula is powerful and straightforward. Unfortunately, for a large n, that is a large number of content items to be considered, this is simply not practical, while too few content items lead to poor results. Since object recognition capability is often suboptimal and computationally extensive, a comprehensive set of content items to represent content is also unlikely to be practical.

On the other hand, domain knowledge can be useful for identifying some content items that bring contextual information. With a relatively small number of content items, this contextual information can be used to obtain a reasonable precision. It helps the classification focus only on contextually relevant images instead of the entire collection. Theoretically, a set of contextually relevant images would consist of relevant images in a reasonably large proportion (compare to that of most of the subsets in the corresponding collection), should the context be correctly chosen. The only problem is how to integrate information from such knowledge and information obtained from data.

In the proposed approach, *Bayesian Networks* (BNs) are used for that purpose. A BN is a directed acyclic graph, where an arc between 2 nodes represents the corresponding causal relationship. BNs enable domain knowledge to be integrated in the network structure in a flexible manner, and help simplify complex joint probability distributions by enabling assumptions to be made on some conditional independence among variables based on some prior knowledge. In particular, they provide a robust framework for incorporating contextual information, which can reduce the need for complete recognition of objects and combat ambiguities.

3.4 An Example

Consider a domain of construction practices commonly found in constructing low-rise commercial buildings in some parts of Indonesia. Such buildings are usually concrete-framed constructed using hand-made timber formwork. Suppose that some construction images related to these practices having the content class "low-rise reinforced concrete building construction" need to be retrieved from a large image collection containing a multitude of images from various domains (such as steel building construction), some of which contain concepts that are not even recognised by knowledge underlying the domain. The retrieval requires classifying these images as having or not having the content class.

Suppose that due to a limited vision capability there are only 3 content items, whose presence or absence can be determined by the corresponding system, that can be used to represent the content class, namely <made_of, form, timber>, <made_of, member, concrete> and <made_of, rebar, steel>. Here, "member" represents either "column" or "beam".

For this example, 350 construction images were studied and processed to compute the required probability values, which can be the basis for ranking the images during the retrieval. These images can be considered relatively good to represent various construction practices from different parts of the world. Let C, c_1, c_2, and c_3 be "low-rise reinforced concrete building construction", <made_of, form, timber>, <made_of, rebar, steel> and <made_of, member, concrete> respectively. The presence and absence of the 3 content items were detected and encoded as 1 or 0 leading to a set of 3 possible values for each of the images.

A retrieval test was then performed on the collection using both direct computation of Eq. (1) and BN-based approach. Two BNs were developed as shown in Figure 3 to represent knowledge concerning relationships among the content items. The first one, representing one expert's knowledge (Figure 3.a), is based on the assumed conditional independence between c_1 and c_2, and between c_1 and c_3, and the second one, representing another expert's knowledge (Figure 3.b), is based on the assumed conditional independence between c_1 and c_2.

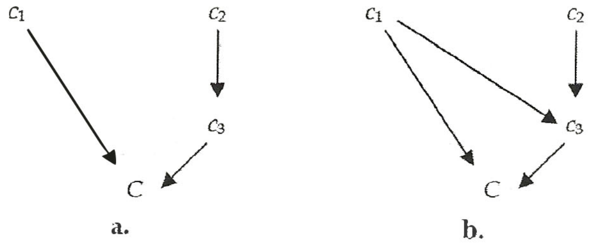


Figure 3. BNs for the Example

From Eq. (1) and applying the conditional independence between c_1 and c_2, and between c_1 and c_3 it can be shown that the first BN results in

$$P(x_4|x_1, x_2, x_3) \cong P(x_4|x_2, x_3)P(x_4|x_1)/P(x_4). \qquad (2)$$

Song *et al* (2002) referred to $P(x_4|x_1)/P(x_4)$ as the context ratio, where x_1 conveys the contextual information. This approach helped the classification focus only on 29 images, which are contextually relevant, out of 350. These are the images that contain $x_1 = 1$.

For the second BN it can be shown that

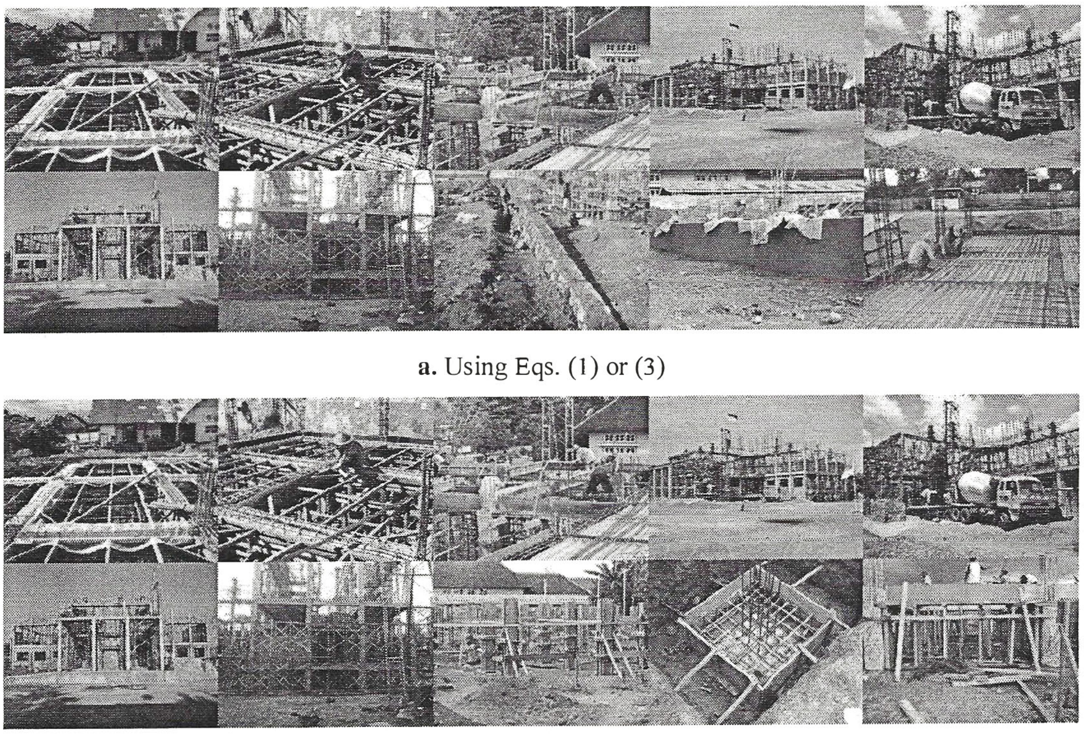
$$P(x_4|x_1, x_2, x_3) \cong \frac{P(x_4|\, x_2, x_3)P(x_4|x_1, x_3)}{P(x_4|\, x_3)} \qquad (3)$$

where the context ratio is $P(x_4|x_1, x_3)/P(x_4|x_3)$. In this test, the number of images to be classified was reduced from 350 to 87. These contextually relevant images contain $x_1 = 1$ or $x_3 = 1$.

The first 10 images retrieved with each approach are shown in Fig. 4. Using Eqs. (1) and (3) resulted in the same first 10 images. Note the absence of any concrete member in the last 3 images retrieved by using Eq. (2). These images were given a high rank since only c_1 (and not c_3) is considered in the context ratio of Eq. (2).

Table 1 shows the retrieval precision for the first 20 images. Here, precision is defined as the proportion of relevant images retrieved. Although the performance of using Eq. (2) is not as good as that of using Eq. (1), it can be improved by modifying the BN structure (Eq. (3)). Comparing the results using Eq. (2) and Eq. (3), the effect of an additional relationship between content items in the modified BN is clear with the increase in precision.

The precision resulting from using Eq. (1) should be the upper bound, since it considers all content items on the entire collection. The use of contextual information reduces the need for such a potentially exhaustive computation, while it is still possible to improve the corresponding performance by modifying the BN structure.



a. Using Eqs. (1) or (3)

b. Using Eq. (2)

Figure 4. The Retrieval Results

Table 1. Precision for the First 20 Images

# Top images	Eq. (1)	Eq. (2)	Eq. (3)
10	1.00	1.00	1.00
15	0.80	0.73	0.80
20	0.60	0.60	0.60

In general, modifying a BN structure can be done by modifying the relationships (either the structure or the probability values) through learning. In this example, however, it is difficult to obtain a reasonably high precision for more than 20 top images, given the 3 content items and only 29 relevant images in the entire collection. One possible way to increase the corresponding upper bound is by adding more defining features, which for this example may include the use of frame scaffoldings or even geographical related features (e.g. sky and soil textures). As the result suggests, the use of BN can handle the fusion of such additional information in a versatile manner.

4. Conclusions

A number of problems related to information retrieval from construction images have been discussed. A framework for a solution is proposed. This is based on the integration of domain-specific concepts and contextual information from domain knowledge along with image content from available data. This paper has made it clear that the fusion of information from both domain knowledge and data using BNs can be useful for construction image understanding. The example demonstrated how it is possible, using this approach, to obtain a reasonable accuracy in construction image classification with a relatively simple computation, given the available content items.

Future work will focus on developing a conceptualisation framework to enable capturing construction domain knowledge for image understanding. This will also include investigating issues on improving the BN structure and developing a learning mechanism.

References

[1] S. J. Simoff and M. L. Maher, "Ontology-based multimedia data mining for design information retrieval", in: K. C. P. Wang, T. Adams, M. L. Maher and A. Songer (eds.), *Proc. Int'l Computing Congress: Computing in Civil Eng.*, 18-21 October 1998, Boston, Massachusetts, US. ASCE, pp. 212-23.

[2] X. B. Song, Y. Abu-Mostafa, J. Sill, H. Kasdan and M. Pavel, "Robust image recognition by fusion of contextual information", *Information Fusion*, **3** (2002), 277-87.

[3] V.-W. Soo, C.-Y. Lee, C.-C. Li, S. L. Chen and C. C. Chen, "Automated semantic annotation and retrieval based on sharable ontology and case-based learning techniques", *Proc. The 3rd ACM/IEEE-CS Joint Conf. on Digital Libraries (JCDL'03)*, 27-31 May 2003, Houston, Texas, US. ACM, pp. 61-72.

[4] J. Fan, Y. Gao, H. Luo and G. Xu, (2005) "Statistical modeling and conceptualization of natural images", *Pattern Recognition*, **38** (2005), 865-85.

[5] J. Luo, A. E. Savakis and A. Singhal, "A Bayesian network-based framework for semantic image understanding", *Pattern Recognition*, **38** (2005), 919-34.

[6] C. B. Tatum, "Classification system for construction technology", *ASCE J. Constr. Eng. & Manag.*, **114** (1988), 344-63.

[7] E. J. Hanlon and V. E. Sanvido, "Constructability information classification scheme", *ASCE J. Constr. Eng. & Manag.*, **121** (1995), 337-45.

[8] N. Maillot, M. Thonnat and A. Boucher, "Towards ontology based cognitive vision", in: J. L. Crowley, J. H. Piater, M. Vincze and L. Paletta (eds.), *Lecture Notes in Computer Science*, **2626** (2003), pp. 45-53.

[9] H. Intraub, "The representation of visual scenes", *Trends in Cogni. Sci.*, **1** (1997), 217-22.

[10] J. M. Henderson and A. Hollingworth, "High-level scene perception", *Annu. Rev. Psychol.*, **50** (1999), 243-71.

[11] L. Hollink, A. Th. Schreiber, B. J. Wielinga and M. Worring, "Classification of user image descriptions", *Int'l. J. Human-Computer Studies*, **61** (2004), 601-26.

[12] L. Wenyin, S. Dumais, Y. Sun, H. J. Zhang, M. Czerwinski, and B. Field, "Semi-automatic image annotation", in: M. Hirose (ed.), *Proc. of INTERACT 2001: 8th TC13 IFIP International Conference on Human-Computer Interaction*, 9-13 July 2001, Tokyo, Japan. IOS Press, Amsterdam, The Netherlands, pp. 326-33.

[13] I. Brilakis, L. Soibelman, and Y. Shinagawa, "Material-based construction site image

retrieval", *ASCE J. Comp. in Civil Eng.*, **19** (2005), 341-55.

[14] H. F. Chua, J. E. Boland and R. E. Nisbett, "Cultural variation in eye movements during scene perception", *Proc. National Acad. Science*, **102** (2005), 12629-33.