The 25th International Symposium
on Automation and Robotics in Construction

June 26–29, 2008

ISARC-2008

Institute of Internet and Intelligent Technologies
Vilnius Gediminas Technical University
Saulėtekio al. 11, 10223 Vilnius, Lithuania
http://www.isarc2008.vgtu.lt/

TWO EFFICIENT TECHNIQUES FOR DIFFICULT APPLICATIONS OF MACHINE VISION IN CONSTRUCTION INDUSTRY

Andrzej Śluzek
Associate Professor
Deputy Director of Robotics Research Centre
School of Computer Engineering, Nanyang Technological University,
Blk N4, Nanyang Avenue, SINGAPORE 639798
assluzek@ntu.edu.sg
tel: +65 6790 4592
fax: +65 6792 6559

ABSTRACT

Machine vision is prospectively the most powerful sensing technique for intelligent automation. However, the existing techniques and algorithms only partially satisfy the needs of construction industry, mainly because of degrading effects that are frequently present at construction sites (weather conditions, muddy, sandy or water-covered sites, partially buried objects, sources of intensive light, etc.). In this paper, we discuss principles of two methods that can significantly improve vision system performances in such conditions (in particular at visually difficult underwater or underground sites). The first method (based on gated imaging) is an efficient image-capturing technique while the second method addresses the problem of object detection in visually complex environments. Both methods have been tested using data and conditions very similar to those expected in real problems, but so far they have not been deployed. The methods can be used either separately or jointly, depending on the problem constraints and hardware availability.

KEYWORDS

Machine vision, object detection, local features, gated imaging.

1. INTRODUCTION

Selected image processing and machine vision algorithms (supported by the corresponding advances in hardware performances) have already reached the level of sophistication and reliability needed in construction industry (or at least in some areas of construction industry). Practitioners have realized that machine vision is prospectively the most powerful sensing technique for intelligent automation. More and more publications report results on machine vision applications in diversified construction-related problems. The range of applications is very wide, but at least some of them can be mentioned to illustrate the potential of machine in construction automation.

Vision can be used to compare the construction progress to the CAD models (e.g. [1]) or to detect faults in the existing structures (e.g. [2] and [3]). Other applications include visual search for database images of construction images (e.g. [4]), vision-based identification of objects in field conditions (e.g. [5]), etc.

The algorithms used in the presented applications are, however, narrowly specialized and heavily exploiting assumptions about the expected contents of the processed images. This disadvantage (unfortunately present in most applications of machine vision) can be attributed to the fact that visual data are almost always highly redundant, contextual and noised. In particular, noised images of unstructured scenes with complex backgrounds (which are typical for construction sites where weather conditions, vegetation, mud, water, alien objects, etc. may distort the quality of images) are very difficult to process and analyze without any preliminary knowledge about their contents.

It is generally believed (and supported by psycho-physiological researches, e.g. [6]) that two key issues towards more universal machine vision techniques are:

(a) Data reduction and enhancement (removing irrelevant data at the earliest stage and/or improving quality of the relevant data).

(b) Detecting data consistency (extraction of visual fragments matching the system needs/know-ledge).

Extensive researches are conducted in these areas and too much space would be required to overview the existing results. However, two illustrative examples can be mentioned. Background removal is a typical data reduction operation in stationary monitoring systems (e.g. [7]) so that only images of intruding object are captured and processed. Visual search for known objects is an example of detecting data consistency. Even if the object is only partially visible, the consistent presence of several local features can be a sufficient evidence of the object's presence. Typical methods of defining, detecting and matching local features can be found, for example, in [8] and [9].

In this paper, two methods are presented that could be particularly useful to construction industry applications of machine vision. The first method, based on *gated imaging*, belongs to (a) group, while the second methods deals with a new category of structure-based local features and their applications in vision algorithms (i.e. the method belongs to (b) category). Although the methods have not been deployed yet in the actual conditions of construction sites, extensive tests have been conducted using images very similar to those expected in real problems.

Sections 2 and 3 of the paper explain principles and show exemplary results of the first method. The second method is presented in Section 4. Section 5 concludes the paper.

2. PRINCIPLES OF GATED IMAGING

Range-gated imaging is a specialized method of acquiring images by using devices that can discriminate between reflected and backscattered light (e.g. [10]). A typical gated imaging system consists of a pulse-generating laser (with the pulses diverged into a conical shape to illuminate a wider area) a high-speed gated camera, and a control and synchronization circuits. Laser-generated pulses reflect from the scene objects and return to a camera equipped with an electronically gated shutter. By synchronizing the gate opening timing with the pulse reflected from objects and returning to the camera, it is possible to capture only the image of those objects.

Reflections from objects that are further away do not return before the shutter closes (i.e. such objects become "invisible") and reflections from closer objects return to the camera before the shutter opens (i.e. these objects are shown as black shadows). Figure 1 illustrates this phenomenon.

By delaying appropriately the gate opening, it is possible to capture an image depicting only those fragments of the observed scene that are within the pre-selected range of distances. Figure 2 shows an image of laser-illuminated scene captured without gating (i.e. with the camera shutter open for a longer period of time) and two gated images for different gate opening delays.

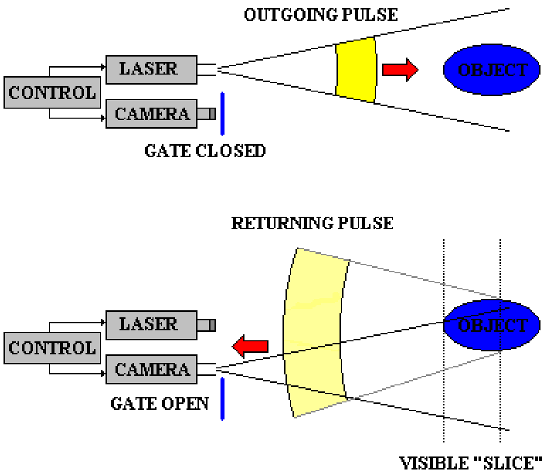


Figure 1. Operational principles of gated imaging

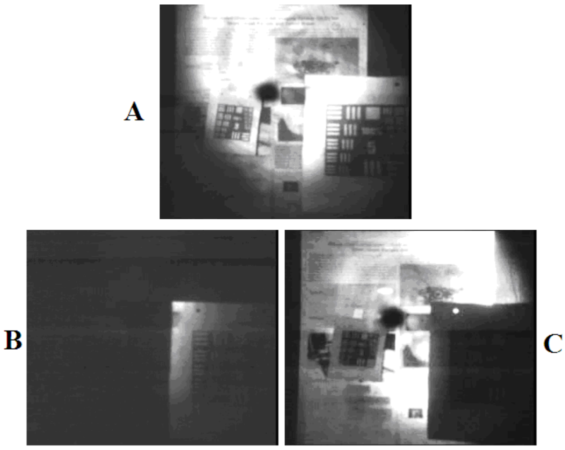


Figure 2. A non-gated image (A) compared to two gated images (B and C) of the same scene

Figure 3 presents a laboratory system used for our experiments with gated imaging techniques. The system produces 10ns pulses (at 10Hz or 20Hz frequency) and the camera-gating time is of similar duration. The system is attached to a 3m long water tank so thank experiment can be conducted both in air and water. More theoretical information on gated imaging can be found in [11] and other papers.

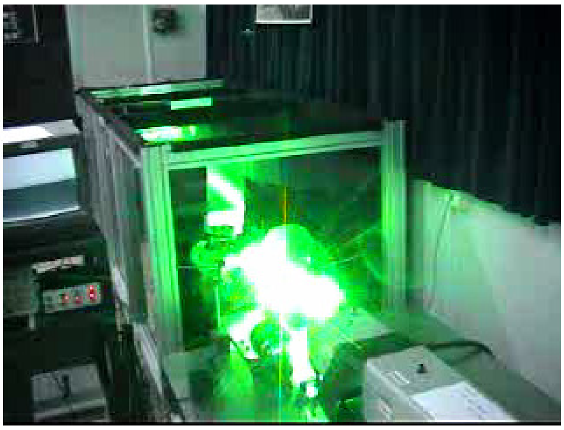


Figure 3. General layout of the gated-imaging equipment

3. OTHER ADVANTAGES OF GATED IMAGING

The most straightforward advantage of gated imaging is the ability to capture objects only within the range of interest. In field applications, the practical implication is that a vision-equipped system would process and analyze only its immediate neighbourhood which is typically the area of interest. Thus, the irrelevant visual information is removed and the amount of data is tremendously reduced. Gated imaging systems, however, can be used in a more sophisticated way because of their additional advantages.

3.1. Turbid-media Imaging

Backscattering noise is an inherent but unwanted component of images captured in turbid media (muddy water, snowfall, heavy rain, etc.). In gated images, the amount of backscattering noise is minimized as most of the backscattered signal (produced by the layer between the camera and the target) returns before the camera gate opens. Generally, gated imaging produces in turbid media images of much better quality (and at longer ranges) than "ordinary" optical devices. It has been verified that the visual penetration of turbid media by gated is 3-6 times deeper than for non non-gated systems.

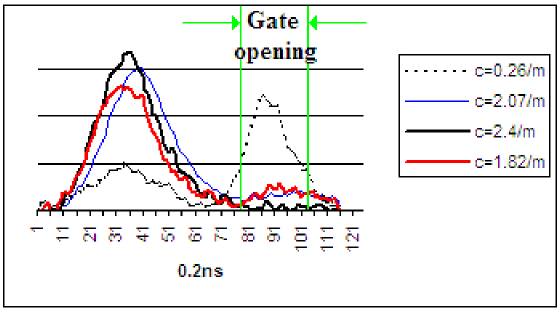
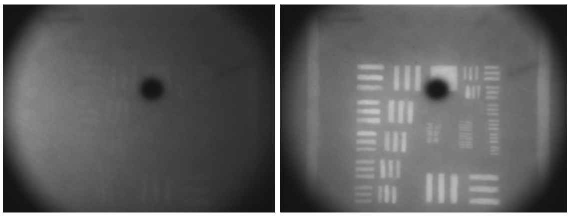


Figure 4. Reflected signal profiles for a 1.8m distant object illuminated by a 10ns laser pulse in water of various turbidities (attenuation *c*)

Figure 4 shows exemplary temporal profiles of a reflected signal in water of diversified turbidity. If the gate opening is correctly delayed, only the object-reflected signal is captured by the camera no matter how much of the backscattering noise is present (though for higher turbidities the signal is weaker). A pair of images (a non-gated *versus* a gated one) captured under the same conditions is compared in Figure 5.



A B

Figure 5. A non-gated image (A) and its gated counterpart (B); both captured in a highly turbid medium

At construction sites where turbid conditions are expected (particularly in underwater construction) a imaging system with such a functionality are obviously an asset.

3.2. Fusion of Gated Images

In realistic scenarios, the scenes can contain several objects at diversified (and possibly unknown) distances. A single gated image may have a too narrow depth to incorporate all those objects (especially in turbid media where the width of "visibility slice" should be narrower to assure a better quality of imaging – see [11]). In such cases, a fusion of several (sufficiently quickly captured) gated images with gradually changing ranges can be used. However, the method would be efficient only if the pieces of visual data are relevantly extracted from individual gated images collected to form the fused image; details can be found in [12].

In Figure 6, three gated images of a complex 3D scene (captured at gradually changing ranges) and the resulting fusion image are compared to a non-gated image obtained under the same conditions.

In actual field applications, fusion of gated images (which requires the adequate camera and laser setting control, processing of individual image and the fused image formation) should be performed online. Though some of these operations can be supervised by a typical computer system, a direct control of system delays needs the timing of nanosecond accuracy. A feasibility-study digital control system has been implemented using an FPGA device as the core component (see [13]). It has been experimentally verified that such a control works reliably and sequences of gated images of gradually changing range can be captured at the fastest rate allowed by the pulsing laser and fused in real time.

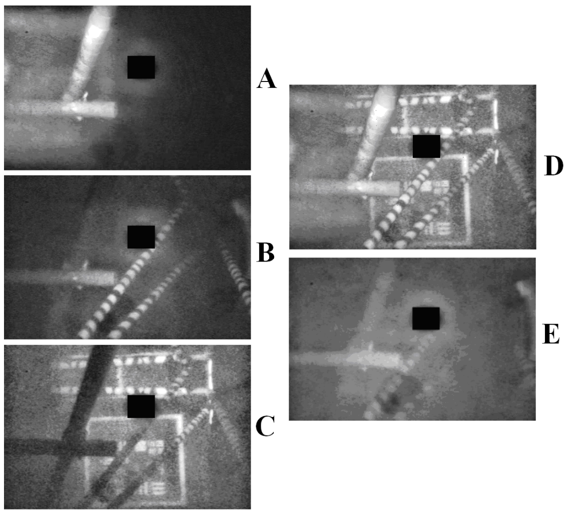


Figure 6. Three gates images captured in turbid water. (A, B, C); (D) – the image fused from gated images; (E) – the corresponding non-gated image shown for comparison

4. LOCAL FEATURES FOR DETECTION OF KNOWN OBJECTS

Several theories exist (e.g. [14]) explaining the human visual perception of objects. The general understanding, however, is that humans recognise known objects as collections of local visual clues which are subsequently (if enough of them are consistently seen) "interpolated" into the object. Thus, machine vision systems (if intended to replace/augment natural vision) should also use local features as the fundamental tool for matching the content of captured images to the models of known objects. Such a mechanism has two advantages. First, it removes redundant/irrelevant data from images (only the features of interest are eventually analyzed) and, secondly, it allows visual detection of known objects under various degrading conditions (occlusions, cluttered scenes, partial visibility due to poor illumination, etc.) when only a part of the objects may be identifiable within images.

In this section, we present the concept of using local features in a more sophisticated way than typical keypoints (e.g. [8], [15]) are used for matching images of similar contents.

4.1. Patterns and Pattern-based Features

The proposed local features (intended for both robotic vision visual information retrieval, see more in [16] and [17]) are produced by approximating fragments of analysed images by selected patterns (defined over circular windows for rotational invariance and for computational efficiency). The patterns of interest are various geometric structures, where each instance of a structure is defined by several geometric parameters and several colours (or intensities). Exemplary instances of selected patterns are shown in Figure 7.

It has been shown that by solving dedicated equations based on low-level moments computed over the circle area, the actual values of the geometric parameters and colours/intensities of the pattern instance can be retrieved (more details in [16] and other previous papers). However, the same equations can be applied to any circular image (not necessarily containing the pattern of interest) of the corresponding radius. If the solutions exist (and it most cases they exist) they specify the *optimum approximation* of that circular image by the pattern of interest. As an illustration, several circular images of various contents are given in Figure 8 together with their pattern-based approximations (using several different patterns). Sometimes, the equations have no solution, i.e. a circular image cannot be approximated by the selected pattern (no such case is shown in Figure 8).

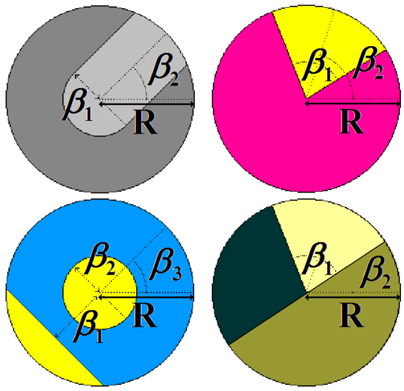


Figure 7. Instances of exemplary circular patterns defined using circles of radius *R* (geometric parameters indicated for each pattern)

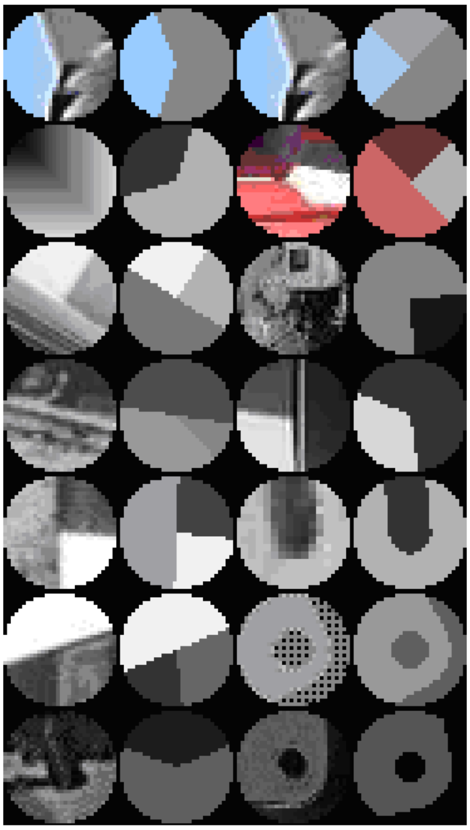
Another straightforward observation is that the visual similarity between a circular and its pattern-based approximation can be diversified. In some cases (when the image actually looks like an instance of the pattern) the similarity is very high, while in other cases there is hardly any similarity.

The similarity between a circular image and its pattern-based approximations can be quantified numerically so that "better" approximations can be automatically

Using the above discussion, we define pattern-based local features as follows:

1. Given any image ***I***, identify locations which are low-level keypoints (e.g. SIFT detector, Harris-Plessey detector – see Figure 9).

2. At each low-level keypoint, place a circular window of a pre-selected radius ***R*** and find its approximations by available patterns.



A B C D

Figure 8. Circular images (A and C) and their approximations (B and D, respectively)

3. If a high-accuracy approximation approximation exists for the location ***K*** for a pattern ***P***, create a candidate pair (***K***, ***P***).

4. For each candidate pair (***K***, ***P***), verify if at the location ***K*** similar (and similarly accurate) approximations exist for the pattern ***P*** if radii longer and shorter than ***R*** are used (i.e. whether the approximation is stable under rescaling).

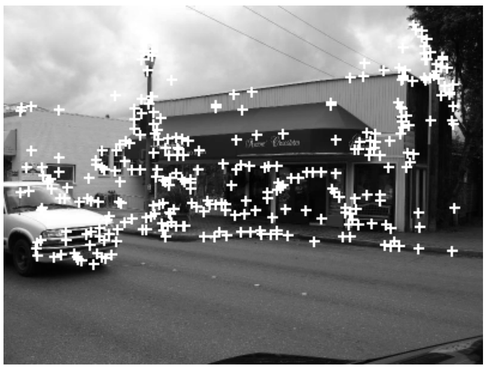
5. A local feature based on the pattern ***P*** exists at any location where an accurate approximation by this pattern is stable under rescaling (within a certain range of scales).

Figure 9 shows a random image and with low-level keypoints detected by Harris-Plessey detector (Fig. 9A) and SIFT detector (Fig. 9B). Although the numbers of keypoints are different in both images, there is a significant number of similar keypoints (especially for visually prominent keypoints) so that the method would search for local features at a sufficient number of the same locations no matter what low-level keypoint detector is used.

Examples of two corner-based local features found in this image are given in Figure 10. For each local feature, two approximations for different radii of the circular windows are presented to show scale-invariance (i.e. stability under rescaling) of the approximations.

It should be noted that the performance of this local feature detection depends on the number of available patterns (in the examples of Figures 9 and 10 only four patterns have been used). With more pattern used, the method becomes more general and suitable for more diversified contents of analysed images.

A



B

Figure 9. Low-level keypoints found in a random image by Haris-Plessey detector (A) and by SIFT detector (B)

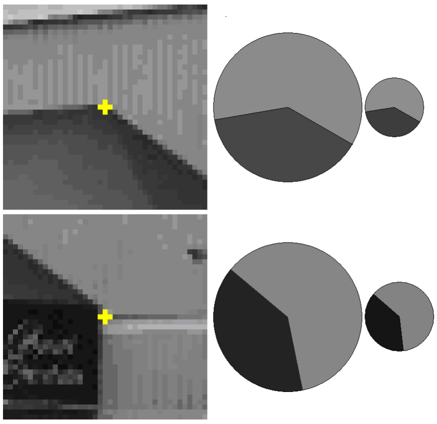


Figure 10. Two low-level keypoints and their corner approximations by (using various radii)

4.2. Prospective Applications of Pattern-based Local Features

Local features based on pattern approximations are very strong, i.e. if exist in an image of an object, they are usually detected in other images of the same object either. Therefore, pattern-based local features are natural candidates for two types of tasks, both prospectively very useful in construction industry.

The first group of applications are in autonomous navigation; in vision-based search for known objects in particular. By analyzing within observed scenes only those local features that are present in the available images of a known object, the autonomous navigator can quickly and effectively identify potential locations of the object even within complex and strongly noised images. More details on how such systems work are available in [9] (although the features used there are not pattern-based).

The second group of applications are in visual information retrieval (which is also an emerging area in construction industry, e.g. [4]). Using pattern-based features detected in query images and in database images, the retrieval can be done efficiently and reliably. Individual matches are strongly confirmed (or rejected) by geometric relations between the features that allegedly belong to the same object. If it is so, all features should be consistently transformed (translated, rotated and their mutual distances should be proportional).

The method works very reliably when objects are partially occluded and seen in changing illumination conditions. However, various experiments have confirmed that matches are possible even if the objects are distorted by linear non-linear transformations (with a limited level of distortion, of course).

Figure 11 shows exemplary pairs of images that were successfully matched using pattern-based features even though the objects of interest are very irregular and/or have partially changed their geometric configurations.

Methods of computationally efficient matching have been presented in many previous works (e.g. [8], [9]) and will not be discussed in this paper.

Figure. 11. Examples of matched images using pattern-based local features

5. CONCLUSIONS

In this paper, two techniques have been presented that are prospectively useful for various machine vision applications in construction industry. Both methods, generally, provide reduction and enhancement of visual data though in two different ways.

In gated imaging, temporal characteristics of light signals acquisition are used to reduce (and control) the visibility depth of observed scenes. Thus, the amount of visual data is reduced by spatial discrimination.

By using pattern-based local feature, the content of captured images is pre-processed in order to detect whether (and where) the image contains visual structures potentially similar to the objects of interest. Thus, the data are reduced by using content-based discrimination.

It is possible to combine both techniques, i.e. the content of gated images may be processed in order to identify pattern-based local features. Such a combination of methods can prospectively become a very powerful machine vision technique for the most difficult conditions (e.g. inside muddy and polluted water reservoirs at construction sites).

Although the methods have not been deployed in natural conditions expected in construction industry, the experiments have been conducted using data fully corresponding to such conditions. In particular, gated images have been captured in a water tank where the level of turbidity can be precisely controlled to match the turbidity of any possibly application.

REFERENCES

[1] Kim, H. & Kano, N. (2008) Comparison of construction photograph and VR image in construction progress, *Automation in Construction*, Vol. 17, 137–143.

[2] Cho, H.T., Jeon, P. & Jung, S. (2004) Implementation and control of a roadway crack tracking mobile robot with force regulation, *Proceedings of the IEEE Int. Conference on Robotics and Automation ICRA 2004*, New Orleans, April 26–May 1, 2244–2249.

[3] Yu S.-N., Jang, J.-H. & Han C.-S. (2006) Auto inspection system using a mobile robot for detecting concrete cracks in a tunnel, *Automation in Construction*, Vol. 16, 255–261.

[4] Brilakis, I. & Soibelman, L. (2005) Content-Based Search Engines for construction image databases, *Automation in Construction*, Vol. 14, 537-550.

[5] Umetani, T. *et al* (2004) Identification of partially-exposed metal object, *Proceedings of the 21st Int. Symposium on Automation and Robotics in Construction*, Jeju, Korea, Sept. 21–25, 274-279.

[6] Biederman, I. (1987) Recognition by components: A theory of human image understanding, *Psychological Review*, Vol. 94, 115–147

[7] Hu, J.-S., Su, T.-M. & Jeng, S.-C. (2006) Robust background subtraction with shadow and highlight removal for indoor surveillance, *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, Beijing, Oct. 9-15, 4545–4550.

[8] Lowe, D. (2004) Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision*, Vol. 60, 91–110.

[9] Islam, M.S. & Sluzek, A. (2008) Relative scale method to locate an object in cluttered environment, *Image and Vision Computing*, Vol. 26, 259–274.

[10] Fournier, G.R., Bonnier, D., Forand, J.L. & Pace, P.W. (1993) Range gated underwater laser imaging system, *Optical Engineering*, Vol. 32, 2185–2189.

[11] Tan, C.S., Sluzek, A. & Seet, G. (2005) Model of gated imaging in turbid medium, *Optical Engineering*, Vol. 44(11), 116002–1–8.

[12] Tan, C.S. & Sluzek, A. (2005) Image formation in highly turbid media by adaptive fusion of gated images, *Lecture Notes in Computer Science*, Vol. 3708, 194–201.

[13] Sluzek, A., Fujishima, H. & Tan, C.S. (2005) Real-time digital control in gated imaging, *Proceedings of 5th EURASIP Conf. on Speech and Image Processing, Multimedia Communication & Services*, Smolenice (Slovakia), June 29 – July 2, 201–206.

[14] Edelman, S. (1997) Computational theories of object recognition, *Trends in Cognitive Sciences*, Vol.1, 298–309.

[15] Schmid, C. & Mohr, R. (1997) Local grayvalue invariants for image retrieval, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, 530–535.

[16] Sluzek, A. (2006) Corner-based keypoints for scale-invariant detection of partially visible objects, *Machine Graphics and Vision*, Vol. 15(3/4), 639–647.

[17] Sluzek, A. (2007) Approximation-based keypoints in colour Images: A tool for building and searching visual databases", *Lecture Notes in Computer Science*, Vol. 4781, 5–16.