A Method for Detection of Known Objects in Visual Search Tasks

Andrzej Sluzek
School of Computer Engineering
Nanyang Technological University
Singapore
assluzek@ntu.edu.sg

***Abstract*—The paper outlines a novel method for visual detection of known objects in images of arbitrary contents. The major improvement over the existing techniques using keypoint detectors (e.g. Harris-Plessey, SIFT) is that the areas around pre-detected keypoints are approximated by various geometric patterns. Matching can be, therefore, done by comparing local geometric and colour structures of keypoints and by testing the geometric consistency over sets of matched keypoint pairs. The method can be used primarily for vision-guided search in mobile autonomous devices. The paper presents the mathematical foundations of the method and illustrates its performances for exemplary problems.**

***Keywords*—keypoints, image matching, geometric patterns, moments, Radon transform**

I. INTRODUCTION

Visual search for objects (for known objects, in particular) is one of the fundamental operations for intelligent agents acting in unknown environments (e.g. exploratory robotics, information retrieval in visual databases, surveillance and security systems, etc.). Even though the physical constraints of the applications might be different (e.g. identification of human faces in a crowded place *versus* search in a digitized database of video-clips) the underlying problems are similar from the perspective of machine vision algorithms. In typical problems, the informational content of visual data is deteriorated and/or distorted by partial occlusions, poor and/or diversified illumination, scale changes, perspective distortions and other similar effects. Under such conditions, the goal is almost always the same, i.e. to detect a rather small fragment from a large amount of irrelevant visual data and to verify whether the extracted fragment matches the visual characteristics of known objects. Usually, the distortions are present both in the analyzed data (e.g. images captured by system cameras) and in the database (i.e. exemplary photos of known objects).

The local approach is a universally accepted machine vision technique for such tasks and this is strongly supported by the evidence (e.g. [1] and [2]) that human vision recognises known objects as collections of local features which are "interpolated" into the object if a sufficient number of those features appears in a consistent configuration. Although this mechanism may not work perfectly (e.g. optical illusions) but generally it allows detection of known objects which are poorly seen because of the degrading conditions listed above.

In machine vision, local features (also referred to as corner points, keypoints, interest points, characteristic points, local visual saliencies, etc.) have been known since the 80's (e.g. [3]). Gradually, they have been used in more sophisticated applications (e.g. motion tracking, image matching and detection of known objects in difficult conditions, e.g. [4]) and currently they are considered a reliable tool for visual information retrieval and vision-guided navigation. SIFT, Harris-Plessey and SUSAN (e.g. [5], [3] and [6]) are examples of the most popular and successful keypoint detectors of low-level local features.

In this paper, we discuss a novel mechanism that uses keypoints as a starting point for detection of known objects. As the numbers of keypoints can be large, image matching becomes a computation-intensive task even with efficient algorithms used, for example, in SIFT detector that produces only highly distinctive features. Generally, keypoints returned by existing detectors are based on differential (or quasi-differential, e.g. SUSAN) properties of the image intensity/colour function. Even though keypoints can be further characterized by various descriptors (e.g. [5] or [7]) those characteristics do not provide any structural description of the corresponding fragments of images. Should such a structural description be available (both in database images of objects and in analyzed images) the search for known objects could be done more selectively and more effectively.

In the proposed method, circular windows are extracted from an analyzed image around the detected keypoints. Each window is subsequently tested whether it contains any of predefined geometric structures. Since only structures present within the database objects are considered, the search immediately focuses on areas possibly containing those objects. Any image fragments of different local geometries are immediately ignored. Finally, the match between database keypoints and image keypoints is based on both structural and geometric consistencies so that presence of known objects can be quickly detected.

The paper presents overview of the method, explains its principles and discusses exemplary results. This is an ongoing research so that the complete evaluation of performances is not available yet. Section 2 presents the algorithm for approximating circular windows by various geometric structures. Accuracy of the approximations is assessed using

the Radon transform (as explained in Section 3). Section 4 illustrates (using selected examples) how the approximations can be used for a quick matching of analyzed images against database images of known objects.

II. Approximations of Circular Images

A. Circular Patterns

Our previous papers (e.g. [8], [9]) described a method that can be find the optimum approximation of an ***R***-radius circular image (grey-level or colour) by a predefined pattern of the same radius. The patterns can be any geometric structure that can be characterized by several configuration and colour/intensity parameters. Although the method is generally applicable to any parameter-defined patterns, corners and corner-like structures (e.g. junctions) are the simplest examples.

Fig. 1 presents several exemplary circular patterns of radius ***R***. As shown in the figure, each instance of a pattern can be uniquely characterised by several geometric (configuration) parameters (represented by indexed βs in Fig. 1) and several colours/intensities. For example, a T-junction (see Fig.1D) can be unambiguously defined by three colours, the orientation angle β_2 and the angular width β_1.

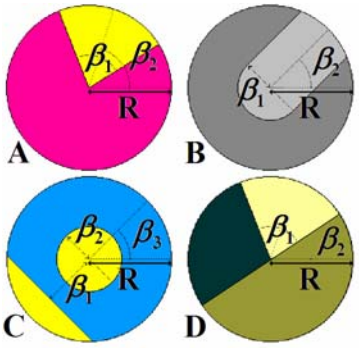

Figure 1. Examplary circular patterns

Various functions of the intensity/colour can be calculated over the pattern area and their analytic forms on the general structure of the pattern while and the function values depend on the parameters of a given instance of the pattern. Moments of the pattern area are examples of such function (they have been proposed as the functions of interest in [8]).

For example, the selected moments of colour corners (Fig.1A) are as follows:

$$\overrightarrow{m_{10}} = \frac{2}{3} R^3 (\overrightarrow{A_1} - \overrightarrow{A_2}) \sin\frac{\beta_1}{2} \cos\beta_2 \qquad (1)$$

$$\overrightarrow{m_{11}} = \frac{R^4}{8} (\overrightarrow{A_1} - \overrightarrow{A_2}) \sin\beta_1 \sin 2\beta_2 \qquad (2)$$

where $\overrightarrow{A_1}$ and $\overrightarrow{A_2}$ are colour vectors (e.g. in RGB space) of both parts of the corner while β_1 and β_2 are angles shown in Fig.1A.

B. Pattern-based Approximations

Given the values of a sufficient number of the area functions of a pattern instance, the configuration and colour/intensities parameters of that pattern can be reconstructed from the set of values. For example, the parameters of patterns shown in Fig.1 can be calculated from of a number of moments (using expressions similar to (1) and (2) to build the corresponding system of equations). As an illustration, selected solutions for configuration parameters for the patterns **A**, **B**, **C** and **D** of Fig.1 are given as follows:

$$\beta_1(A) = 2\arcsin\sqrt{1 - \frac{16\left(\left\|\overrightarrow{m_{20}} - \overrightarrow{m_{02}}\right\|^2 + 4\left\|m_{11}\right\|^2\right)}{9R^2\left(\left\|m_{10}\right\|^2 + \left\|m_{01}\right\|^2\right)}} \qquad (3)$$

$$\beta_2(B) = \operatorname{atan2}(\pm m_{01}, \pm m_{10}) \qquad (4)$$

$$\beta_1(C) = \sqrt{\frac{\left(\overrightarrow{m_{20}} - \overrightarrow{m_{02}}\right)^2 + 4\left\|\overrightarrow{m_{11}}\right\|^2}{\left(\left\|\overrightarrow{m_{10}}\right\|^2 + \left\|\overrightarrow{m_{01}}\right\|^2\right)}} \qquad (5)$$

$$m_{01}\cos\beta_2(D) - m_{10}\sin\beta_2(D) = \pm\frac{4}{3R}\sqrt{(m_{20} - m_{02})^2 + 4m_{11}{}^2} \qquad (6)$$

It should be noted that some solutions are for colour images for which moments are 3D vectors in RGB space, and some for grey-level images where moments are scalars. Intensity/colour parameters of the pattern instances can be found using other moment-based formulae (also individually designed for each pattern). Examples can be found in [8] and [9].

If the number of parameter increases (i.e. if the patterns become more complex) a larger number of moments should be used. This is not recommended since only low orders are sufficiently insensitive to noise. Therefore, moments of the pattern halves and quarters can be used instead so that 54 moment expressions exist using moments of order 0, 1 and 2 only (six moments of the whole circle, six moments of the upper half, six moments of the first quarter, etc. – more details are given in [10]).

Equations such as (3) – (6) can be actually applied to any circular image not necessarily containing the corresponding pattern. However, if the solutions exist for a circular image of arbitrary content, they define the *optimum approximation* of that image by the given pattern. In some cases solutions do not exist (e.g. a negative number in (3)) which means that the circular image cannot be approximated by a given pattern.

Examples of circular images and their approximations by selected patterns are given in Fig. 2. It can be seen that for images containing actual patterns the corresponding approximations look very similar. However, there are also examples of approximations looking very different than the images. In some cases, the results are ambiguous. If the image content is inconclusive it may me approximated by various

patterns with visually similar accuracy (e.g. the first row in the figure).

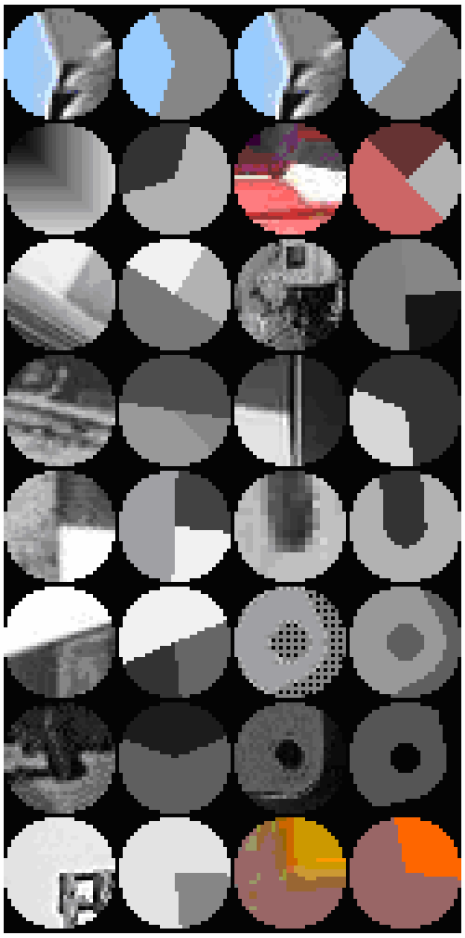

Figure 2. Circular images (colums 1 and 3) and their approximations (columns 2 and 4, respectively) by exemplary patterns.

III. MATCHING PATTERN-BASED APPROXIMATIONS

Pattern-base approximations are the proposed technique for matching processed images to the images of known objects. Given a set of pre-detected keypoints (using any sufficiently reliable keypoint detector) we can build for each keypoint approximations of the surrounding circular area for any available patterns. Since the same operation would be performed both for database keypoints and for keypoints extracted from analyzed images. If for a subset of database keypoint and image keypoint the approximations are consistent (both for individual pairs of matched keypoints and over the whole subset of keypoints) there is a strong indicator that the object is present within the image. However, as shown in Fig.2, the approximations may actually look very different than the approximated circular image. Obviously, the matches should be based only on those keypoints for which the approximations accurately correspond to the approximated circular areas around the keypoints.. Therefore, a measure is needed to compare a circular image and its pattern-based approximation.

A. Similarity Measuress

Several measures have been proposed in our previous paper (e.g. based on differences between the circular image and its computer-synthesized approximation, [8] or based on moment similarities, [9]). After extensive experimental verification it was found that the most reliable results are obtained using a well-known Radon transform (e.g. [11]) calculated for a few directions. Actually, sufficiently that satisfactory results can be achieved using the Radon transform computed just for the horizontal and vertical directions.

The formulae for the horizontal and vertical Radon transforms can be derived analytically for any pattern of interest. For example, for a grey-level corner the Radon transform computed along horizontal lines can be expressed as

$$F(y)=\begin{cases} 2A_2\sqrt{R^2-y^2} \text{ if } |y|>R\sin\dfrac{\beta_1}{2} \\ (A_1+A_2)\sqrt{R^2-y^2}+(A_2-A_1)|y|\,ctg\dfrac{\alpha}{2} \text{ if } |y|\le R\sin\dfrac{\beta_1}{2} \end{cases} \quad (7)$$

where A_1 and A_2 are intensities of both parts of the corner. Note that the orientation angle β_2 would be the offset value for calculating the Radon transform for the approximated circular image.

Fig.3 shows examples of an accurate (Fig.3A), poor (Fig.3B) and average (Fig.3C) approximation of circular images by corners. Selected corresponding diagrams (illustrating differences between Radon transforms of images and their approximations) are presented in Figs.4-6.

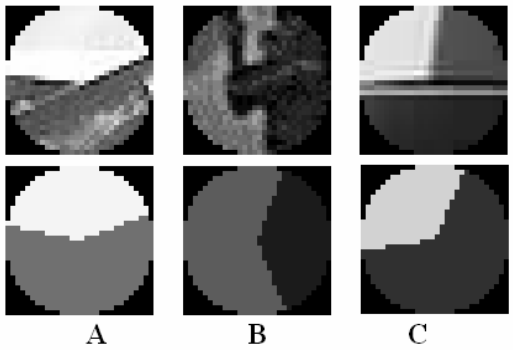

Figure 3. Three circular images and their corner approximations of good (A), poor (B) and average (C) quality.

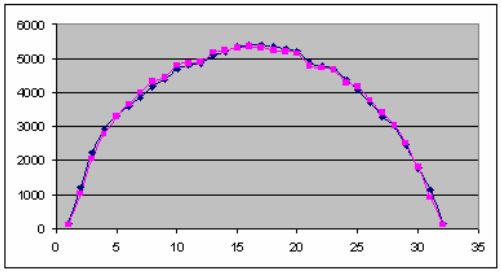

Figure 4. Vertical Radon transforms for Fig.3A (blue: circular image, pink: its corner approximation).

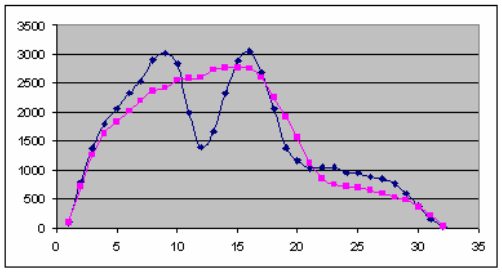

Figure 5. Vertical Radon transforms for Fig.3B (blue: circular image, pink: its corner approximation).

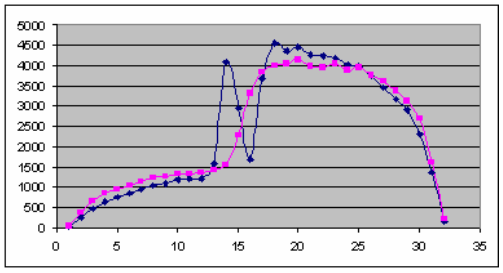

Figure 6. Horizontal Radon transforms for Fig.3C (blue: circular image, pink: its corner approximation).

The integrated difference between horizontal and vertical Radon transforms is used as the measure of the approximation accuracy. Additionally, the measure is further normalized using the scaling factor depending on the intensity of approximations and contrasts between various part of the approximations. Without the normalization, dark and/or poorly contrasted images would be (incorrectly) too similar to their correspondingly dark and/or poorly contrasted approximations. Altogether, the similarity between ***I*** circular image and its ***App*** approximation (both of radius ***R***) can be expressed as

$$Sim(I, App) = \frac{K}{R^2\left[1 + diff(I, App)\right]}$$

where

$$diff(I, App) = \int_{-R}^{R} \left| Rad_I(0^\circ, s) - Rad_{App}(0^\circ, s) \right| + \left| Rad_I(90^\circ, s) - Rad_{App}(90^\circ, s) \right| ds \quad (8)$$

B. *Principles of matching*

In the proposed method, pattern-based approximations are computed for circular areas around pre-detected keypoints. Various keypoint detectors can be used, but within this paper we use only Harris-Plessey and SIFT examples. The important issue is the radius selection for the circular area around the keypoints. Generally, the radius size should correspond to the scale of detectors. In the Harris-Plessey detector the scale is uniform for all keypoints (determined by the standard deviation of the Gaussian convolution used) so that the same radius should be used for all keypoints. Alternatively, SIFT returns the optimum scale for each detected keypoint. In this case, the radius should be individually determined for each keypoint proportionally to its scale.

Fig.7 shows an exemplary image (a smooth image has been deliberatly selected) with keypoints detected by both Harris-Plessey or SIFT operators. Although the keypoints are not identical, their general distribution is very similar and, more importantly, numerous visually prominent keypoints are identical or almost identical.

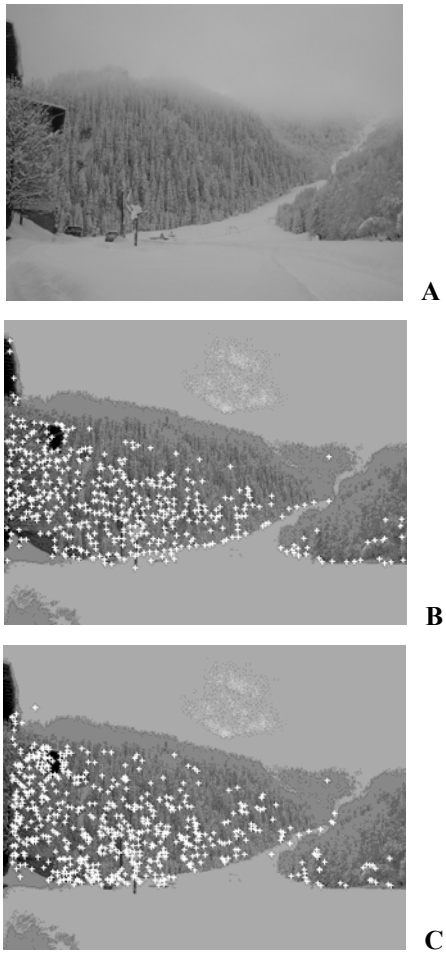

Figure 7. Exemplary image (A) with keypoints detected by Harris-Plessey (B) and SIFT (C) operators.

Fig.7 and many other similar examples convincingly indicate that a sufficient number of (almost) identical keypoints are available for pattern-based approximation no matter what keypoint detector is used.

Using the proposed method, the database keypoints are matched to the keypoints extracted from analysed images by comparing their pattern-based approximations over circular areas. This approach has several significant advantages:

- Keypoints can be categorized into several groups based on which approximations are the most accurate for each keypoint.
- If the database keypoint are accurately approximated by a limited number of patterns, only those image keypoints that are accurately approximated by the same pattern should be considered..

- Parameters of the approximations can be very selectively used in the matching process. For example, the colour/intensity parameters may be used explicitly (matched keypoint should have similar values) may be used selectively (e.g. using "darker" or "lighter" relations only) or may be totally ignored (so that geometry of the patterns should be similar but any colours/intensities are accepted).

IV. EXAMPLES

To illustrate performances of approximation-based keypoint matching, two examples are presented in this section. Fig.8 shows two different views of the same object with keypoints detected by Harris-Plessey operator. Selected keypoints representing the same parts of the object are (manually) matched as shown in the figure.

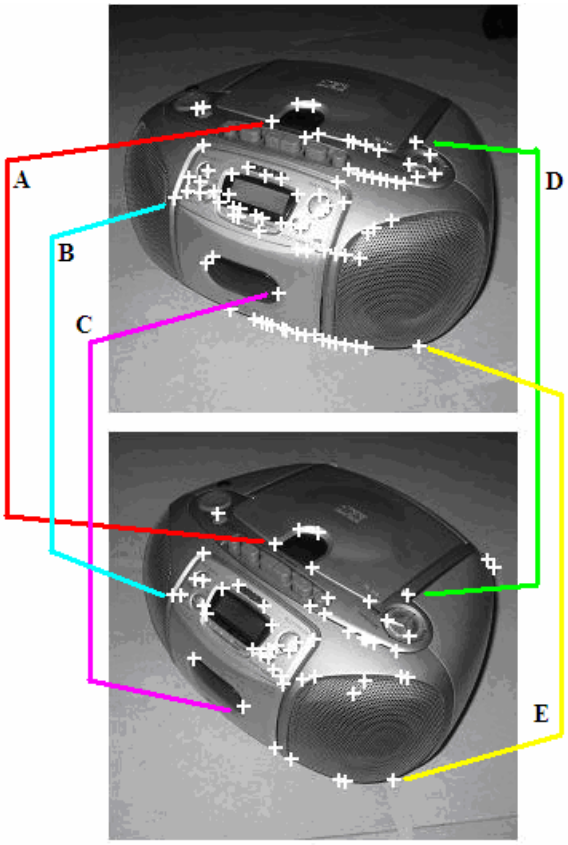

Figure 8. Exemplary pairs of matched keypoints for two views of the same object.

Table 1 contains geometric and intensity parameters for the pattern-based approximations of the matched pairs. In this example only four types of patterns have been used (namely: *corners*, *orthogonal T-junctions*, *sectional cuts* and *X-crossings*). Although details of the image matching techniques are not discussed in this paper (more can be found in [7] and other publications) the results given in the table – in particular consistent orientation differences between approximations of matched pairs –strongly indicate that all keypoints may belong to the same object. Additionally, the relative transformation of the object between both images can be roughly estimated from the spatial distribution of the keypoints and from their geometric parameters.

TABLE I. APPROXIMATION PARAMETERS FOR FIG.8 KEYPOINTS

Pair A	*Pattern*: **90° T-junction**		
lower image	Intensities: A = 66 B = 98 C = 33	Orientation: β = 65°	
upper image	Intensities: A = 55 B = 100 C = 41	Orientation: β = 59°	
Pair B	*Pattern*: **Corner**		
lower image	Intensities: A = 127 B = 69	Orientation: β = 3°	Angular width: α = 141°
upper image	Intensities: A = 143 B = 69	Orientation: β = -5°	Angular width: α = 150°
Pair C	*Pattern*: **X-cross junction**		
lower image	Intensities: A = 92 B = 129	Orientation: β = 81°	
upper image	Intensities: A = 88 B = 127	Orientation: β = 68°	
Pair D	*Pattern*: **Corner**		
lower image	Intensities: A = 115 B = 70	Orientation: β = 92°	Angular width: α = 81°
upper image	Intensities: A = 95 B = 56	Orientation: β = 84°	Angular width: α = 88°
Pair E	*Pattern*: **Corner**		
lower image	Intensities: A = 78 B = 168	Orientation: β = -103°	Angular width: α = 174°
upper image	Intensities: A = 77 B = 162	Orientation: β = -109°	Angular width: α = 173°

Fig. 9 shows a selected pair of SIFT-detected keypoints (with the approximated circular areas around) for two images of an animal.

Figure 9. Two images of an animal with a pair of matched keypoints.

TABLE II. APPROXIMATION PARAMETERS FOR FIG.9 KEYPOINTS

	Pattern: **90° T-junction**		
left image	Intensities: A = 128 B = 79 C = 28	Orientation: β = -21°	
rigth image	Intensities: A = 125 B = 80 C = 25	Orientation: β = -33°	
	Pattern: **Corner**		
left image	Intensities: A = 36 B = 128	Orientation: β = -54°	Angular width: α = 146°
rigth image	Intensities: A = 31 B = 127	Orientation: β = -70°	Angular width: α = 145°
	Pattern: **X-cross junction**		
left image	Intensities: A = 78 B = 103	Orientation: β = -21°	
rigth image	Intensities: A = 75 B = 102	Orientation: β = -33°	
	Pattern: **Sectional cut**		
left image	Intensities: A = 30 B = 125	Orientation: β = -54°	Cut ratio: t = 0.78
rigth image	Intensities: A = 25 B = 124	Orientation: β = -70°	Cut ratio: t = 0.77

The content of Table 2 confirms that approximations are almost identical (for all four patterns). The only significant difference is the orientation angle but it consistently differs by approx. 15° for all patterns.

V. CONCLUDING REMARKS

The paper presents an overview and preliminary results of a novel technique of defining local features. The features are based on keypoints extracted by any typical detector but subsequently circular areas around the keypoints are approximated by selected patterns. Eventually, only those keypoints are used for image matching where the areas (i.e. the local image content) can be satisfactorily accurately approximated by at least some patterns.

The method could be particularly useful for autonomous navigation where agents exploring unknown environments should locate known objects and/or should characterize the observed scene in terms of its similarity to already known environments. Visual information retrieval (VIR) where pictorial databases are searched for images containing fragments similar to known objects is the second intended application.

The paper discusses only the fundamentals of the method, i.e. the problem formulation, basic definitions and the mathematical principles of pattern-based approximations. Examples are included to explain and illustrate those fundamentals.

REFERENCES

[1] I. Biederman, "Recognition-by-components: A theory of human image understanding", Psychological Review, vol. 94(2), pp. 115-147, 1987.

[2] S. Edelman, "Computational theories of object recognition", Trends in Cognitive Sciences, vol. 1(8), pp. 298-309, 1997.

[3] C.Harris, and M.Stephens, "A combined corner and edge detector", Proc. of 4th Alvey Vision Conference, pp 147-151, 1988.

[4] C.Schmid, and R.Mohr, "Local grayvalue invariants for image retrieval", IEEE Trans. PAMI, vol. 24(5), pp. 530-535, 1997.

[5] D. Lowe, "Distinctive image features from scale-invariant keypoints", Int. Journal of Computer Vision, vol. 60(2), pp 91-110, 2004.

[6] S.M. Smith, and M.Brady, "SUSAN – a new approach to low level image processing". Int. Journal of Computer Vision, vol. 23(1), pp. 45-78, 1997.

[7] M.S.Islam, and A.Sluzek, "Relative scale method to locate an object in cluttered environment", Image & Vision Computing, vol. 26(2), pp. 259-274, 2008.

[8] A. Sluzek, "On moment-based local operators for detecting image patterns", Image and Vision Computing, vol. 23(3), pp 287-298, 2005.

[9] A.Sluzek, "Approximation-based keypoints in colour images: A tool for building and searching visual databases", *LNCS*, Vol.4781 (eds G.Qiu, C.Leung, X-Y.Xue, R.Laurini), Springer Verlag, 2007, pp 5-16.

[10] A.Sluzek, "Keypatches: A new type of local features for image matching and retrieval", Proc. of 16th WSCG Conference, 2008 (in print)

[11] S.R. Deans, The Radon Transform and Some of Its Applications, John Wiley & Sons, New York, 1983.