We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

New Types of Keypoints for Detecting Known Objects in Visual Search Tasks

Andrzej Śluzek[1,2] and Md Saiful Islam[1,3]
[1]Nanyang Technological University, [2]SWPS, [3]Dhaka University of Engineering and Technology
[1]Singapore, [2]Poland, [3]Bangladesh

1. Introduction

Visual exploration of unknown environments is considered a typical and highly important task in intelligent robotics. Although robots with visual capabilities comparable to human skills (e.g. mushroom-picking robots or bird-viewing robots) are apparently unachievable in the near future, but the concept of robots able to search for known objects in unknown surroundings is one of the ultimate goals for machine vision applications. In the scenarios that are currently envisaged, the expectations should be realistically limited. Nevertheless, one can expect that a robot, after a visual presentation of an object of interest, should be able to "learn" it and, subsequently, to detect the same object in complex scenes which may be degraded by typical effects, i.e. partial visibility of the objects (due to occlusions and/or poor illumination) and their unpredictable locations. The purpose of this chapter is to propose a novel mechanism that is potentially useful (it has been confirmed by promising preliminary results) in such applications.

Several theories exist explaining the human perception of objects (e.g. Edelman, 1997). Some researchers promote the importance of multiple model views (e.g. Tarr et al., 1997) others (e.g. Biederman, 1987) postulate viewpoint invariants in the form of shape primitives (geons). From all the theories, however, the practical conclusion is that vision systems detecting objects in a human-like manner should use locally-perceived features as the fundamental tool for matching the scene content to the models of known objects.

The idea of using local features (keypoints, local visual saliencies, interest points, characteristic points, corner points – several almost equivalent names exist) in machine vision can be traced back to the 80's (e.g. Moravec, 1983; Harris & Stephens, 1988). Although stereovision and motion tracking were initially the most typical applications, it was later found that the same approach can be used in more challenging tasks (e.g. matching images in order to detect partially hidden objects). A well-known Harris-Plessey operator (Harris & Stephens, 1988) was combined with local descriptors of detected points to solve object recognition problems in which local features from analysed images are matched against a database of images depicting known objects (e.g. Schmid & Mohr, 1995). The intention was to retrieve images containing arbitrarily rotated and partially occluded objects.

Subsequently developed keypoint detectors address the issues of scale changes (this was the weakest point of the original detectors) and perspective distortions. Generally, to achieve

scale invariance of local features, computationally expensive scale-space approaches are used (e.g. Lowe, 2004; Mikolajczyk & Schmid, 2004). So far, no method is known that can scale-invariantly match local features using a one-size window for scanning images captured in arbitrarily changing scales. The perspective distortions are usually approximated by affine transformations (or even ignored altogether). This is acceptable since only relatively minor distortions are typically assumed. Stronger deformations are avoided by using multiple views (differing usually by 15-30 degrees) to model 3D database objects.

Our paper presents how to integrate and expand selected ideas from the abovementioned theories and techniques into an alternative framework that could satisfy the practical requirements of robotic vision systems at lower computational costs than other currently existing solutions. Generally, we follow the fundamental concepts presented in previously published works (e.g. Huttenlocher & Ullman, 1990; Wolfson & Rigoutsos, 1997; Häusler & Ritter, 1999; Ulrich et al., 2003; etc.). In particular:

1. Database objects are represented by 2D images. Multiple images of the same object are used if 3D transformations of the object are expected in the captured scenes, while a single image is needed if only 2D transformations are envisaged.
2. Database objects are modelled as a set of locally computed features (keypoints) characterised by their descriptors. The geometric constraints of the set (i.e. length and orientation of vectors joining the keypoints) are also stored.
3. Keypoints of the same categories are extracted from captured scenes. Subsequently, those keypoints are matched to the models of database objects. If a sufficient number of the keypoints are consistently (i.e. satisfying the geometric constraints of the model) matched to a certain model image, the corresponding database object is considered found in the scene.

What makes our method novel is the definition of local features (keypoints). Therefore, the major sections of this chapter discuss the proposed keypoints and present exemplary results obtained using such keypoints. The actual object detection and/or localisation are only briefly mentioned since the methods used follow the algorithms published in our previous papers or papers of other authors.

Typically used keypoints are based directly (e.g. Harris & Stephens, 1988) or indirectly (e.g. Lowe, 2004; Mikolajczyk & Schmid, 2004) on derivatives of the intensity functions. Such keypoints have many advantages but certain disadvantages as well. For example, the scanning window over which the keypoints are computed should be resized according to the scale of objects present in the image. If the scale is unknown (which is the most typical scenario) additional computations and/or assumptions are necessary. Some authors use computationally intensive search for the optimum scale at which the current keypoint should be processed (e.g. SIFT detector in Lowe, 2004) while others propose a simplifying (but nevertheless justifiable for robotic application) assumption that only a few scales are used and the object would be identified when its distance to the capturing camera corresponds to one of those scales (e.g. Islam et al., 2005). An additional disadvantage of derivation-based keypoints is that some photometric transformations (e.g. excessively high contrasts) may distort the captured image to the point where the original differential properties of the intensities are lost while the visual content of the image is still readable.

We propose keypoints based on the local structural properties of the images, i.e. the contents of scanning windows are approximated by a certain number of structures (parameterised

patterns). If the approximation if sufficiently accurate, a keypoint is built and characterised by the parameters describing this approximation. The fundamentals of such keypoints are presented in Section 2.

In Section 3, we discuss how to use such approximation-based keypoints for object detection (including scale-invariance issues). It is shown that, in spite of using uniform scanning windows, objects at arbitrary scales can be matched (within a certain range of scales). Section 4 presents exemplary results of the proposed technique and briefly explains the further steps of object detection. Conclusions and additional remarks are given in Section 5.

2. Approximation-based Keypoints

2.1 Pattern-based Approximations

Recently (in Sluzek, 2005) a method has been proposed for approximating circular images with selected predefined patterns. Although corners and corner-like patterns (e.g. junctions) are particularly popular and important, the method is applicable to any parameter-defined patterns (both grey-level and colour ones, though the latter are not discussed in this chapter).

We assume that a *grey-level circular pattern* is modelled by several configuration parameters and intensity parameters (as shown in exemplary patterns given in Figure 1). Typical patterns are specified by 2-3 configuration parameters and 2-3 intensities. The radius R of a pattern can be arbitrarily selected. Thus, if a configuration parameter is a length (e.g. β_1 in Figure 1B, or β_1 and β_2 in Figure 1C) it should be measured both absolutely and relatively to the radius.

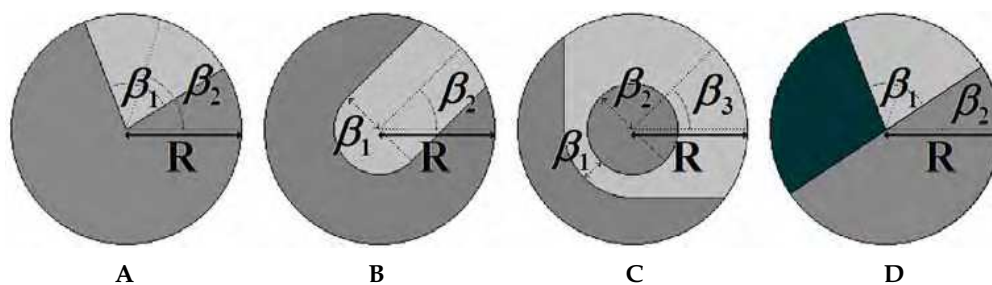A   B   C   D

Figure 1. Exemplary patterns defined in circles of radius R (configuration parameters shown)

Circular patterns are considered templates that would be matched to other circular images (or rather to circular windows of a larger image) in order to determine how well that image can be approximated by given patterns. In other words, the optimum values of the parameters should be found to identify the best pattern approximation. This idea (originally applied to edge detection) can be traced back to the 70's (e.g. Hueckel, 1973).

In our previous papers (e.g. Sluzek, 2005) it is explained how to build the optimum approximations for various template patterns (or, alternatively, how to determine that no such approximation exists) using locally computed intensity moments. For several patterns, the explicit solutions are given. For example, the orientation angle β_2 of a corner approximation (see Figure 1A) is obtained from

$$\beta_2 = \arctan 2(\pm m_{01}, \pm m_{10}) \tag{1}$$

while the angular width β_1 is computed as

$$\beta_1 = 2\arcsin\sqrt{1 - \frac{16[(m_{20} - m_{02})^2 + 4m_{11}^2]}{9R^2\left(m_{10}^2 + m_{01}^2\right)}} \tag{2}$$

For T-junctions (Figure 1D) β_1 angular width and β_2 orientation angle can found from

$$\frac{\pi}{2} - \beta_2 - \frac{\beta_1}{2} = \frac{\arctan 2(\pm m_{02} \mp m_{20}, \pm 2m_{11})}{2} \tag{3}$$

and

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

where m_{10}, m_{20}, etc. are moments of the corresponding orders computed in the coordinate system attached to the centre of circular windows.

The intensities of such approximations can be estimated using other moment-based expressions (details in Sluzek, 2005).

Exemplary circular windows (containing actual corners, T-junctions and more random contents) are given in the top row of Figure 2. The bottom row shows the optimum corner or T-junction approximations. For some irregular images the approximations do not exist, i.e. the corresponding equations have no solutions.
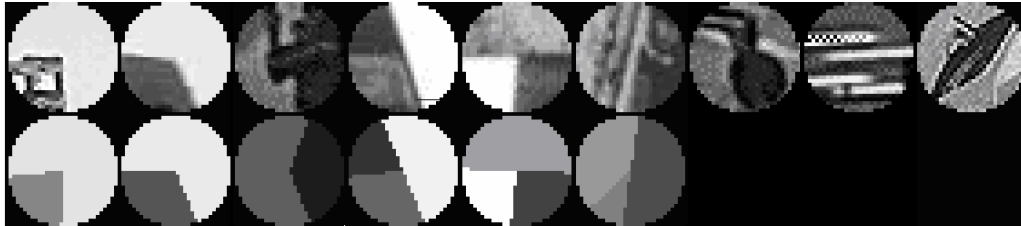

Figure 2. Optimum approximations (using corner or T-junction patterns) for selected circular images of 15-pixel radius

It can be straightforwardly proven that results produced by Eqs (1)-(4) are invariant under linear illumination changes, and that non-orientation configuration parameters (e.g. angular widths β_1 in Figs 1A and 1D) are invariant under any 2D similarity transformation. Extensive tests have also indicated that the results are stable (unlike, for example, the corner approximations discussed by Rosin, 1999) under high- and low-frequency noise, image texturization and partial over- and under-saturation of intensities. The same level of stability has been confirmed for other circular patterns.

2.2 Approximation-based Model Keypoints

Examples in Figure 2 show that even if the approximation exists, there might be a significant visual difference between a circular image and its approximation. Thus, if we can measure the level of similarity between an image and its approximation, the optimum approximation

(i.e. the approximation with the highest similarity) indicates how accurately the pattern of interest is actually "seen" in the image.

Alternative methods of quantifying similarity between an image and its pattern approximation have been given in past papers (Sluzek, 2005; Sluzek, 2006). Unfortunately, the complexity of both methods is as high as the complexity of building the approximations. It has been eventually found that highly satisfactory results can be achieved in a simpler way by comparing moments of circular images (these moment have to be computed anyway) and moments of their approximations (those moments can be immediately calculated from the parameters of the approximation). Thus, the similarity between a circular image I and its approximation AI can be quantified using one of the following *similarity functions*:

$$sim_1(I, AI) = K - \alpha \frac{abs(m_{20} - ma_{20}) + abs(m_{02} - ma_{02}) + abs(m_{11} - ma_{11})}{m_{20} + m_{02} + abs(m_{11})} \qquad (5)$$

$$sim_2(I, AI) = K - \alpha \frac{abs(m_{10} - ma_{10}) + abs(m_{01} - ma_{01})}{abs(m_{10}) + abs(m_{01})} \qquad (6)$$

where m_{pq} and ma_{pq} are moments of I and AI (respectively) and K, α are arbitrarily selected positive values.
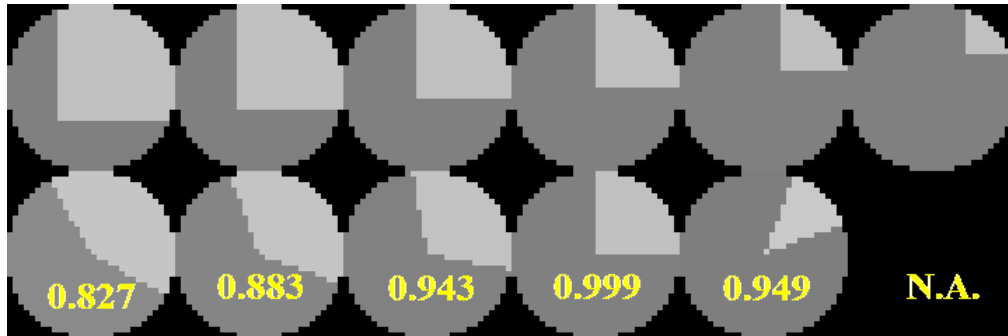

Figure 3. *Top row*: a sequence of windows moving across a high-quality corner image. *Bottom row*: corresponding corner approximations and the similarity levels (for the last window the corner approximation does not exist)

If at certain location an image contains a fragment similar to the pattern of interest, a high level of similarity between the content of a scanning window located there and its approximation is expected. However, a high similarity level would be found not only for the actual location but also for neighbouring locations. The similarity, nevertheless, reaches a local maximum at the location. Figure 3 illustrates this effect.

Moreover, if an image contains a certain pattern, the similarity between the window content and the approximations exists for a certain range of radii of the scanning window and the approximations are consistent over this range of radii (instead, the scanning window may remain the same, but the image is resized correspondingly over the range of scales). An example showing such a consistency both for the configuration and intensity parameters for a selected fragment of a digital image (containing a T-junction) is given in Figure 4.

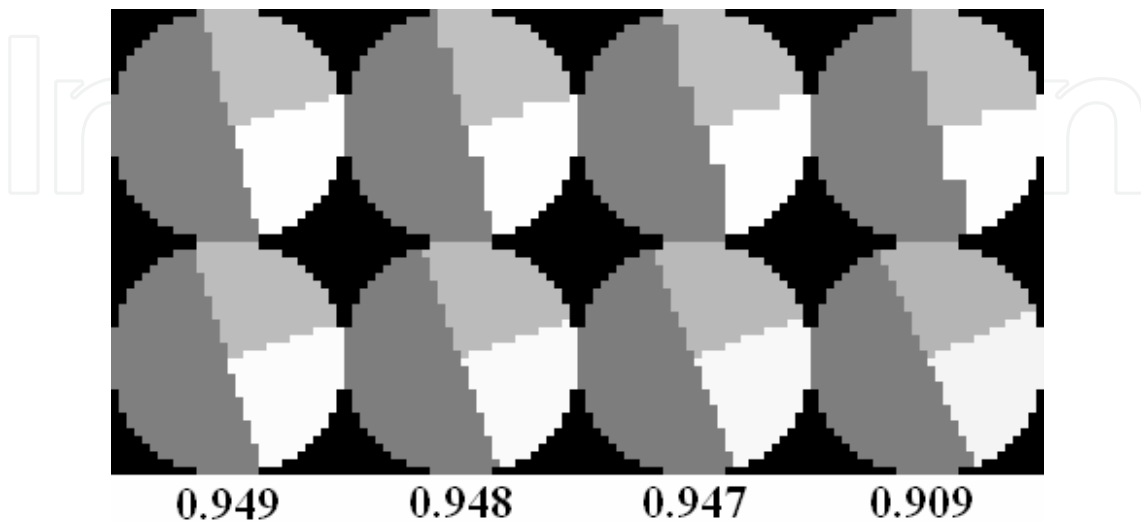Thus, our proposal of the novel type of keypoints is based on the above discussion.

Figure 4. *Top row*: a sequence of 15-pixel windows over a gradually enlarged image of a T-junction. *Bottom row*: corresponding T-junction approximations and the similarity levels

Definition 1

For a given image, pixel (x,y) is (subject to additional requirements explained below) an *approximation-based model keypoint* (shortly *model keypoint*) defined by a circular pattern **TP** if for the scanning windows located at (x,y):

1.   **TP** pattern-based approximation exists for each radius **R** from a certain range (R_1, R_2).
2.   The approximations have consistently similar parameters over the whole range of radii (R_1, R_2).
3.   If several neighbouring pixels satisfying (1) and (2) exist, the model keypoint is located at the pixel where the *similarity* between the scanning windows and their approximations reaches a local maximum.

Typically recommended additional requirements (introduced for practical reasons) are as follows:

•   Similarities between the window contents and the approximations should be sufficiently high (keypoints that inaccurately depict the pattern are less useful than the accurate ones).

•   Contrasts between intensity parameters of the produced approximation (see Figure 1) should exceed a predefined threshold (keypoints that can be hardly seen usually have no practical importance). For 256-level images, the recommended thresholds are in 15-25 range.

•   The similarity functions can by additionally modified proportionally to the contrasts between intensities of the produced approximation (less accurate keypoints with better contrasts might be more important than poorly contrasted keypoints of high accuracy).

•   Pattern-specific constraints may exist. For example, the angular width of a corner approximation should not be too close to 180° (it becomes an edge then) or to 0° (it effectively becomes a line tip).

It should be noted that the proposed definition of approximation-based model keypoints is not limited by the proposed method of computing the approximations. In fact, the definition is applicable to any other technique where image fragments similar to selected patterns of interest are searched for.

We propose to use the above-defined *model* keypoints for *model* images of the objects of interest. First of all, such keypoints are stable prominent features that are likely to be preserved in any other image that contains the same fragmnent of the object even if the viewing conditions are changed. Secondly, the number of such high-quality keypoints is usually limited (for a single pattern) even in complex objects. However, if several different patterns are used, the model image can still contain enough keypoints for a reliable detection under partial occlusions. Nevertheless, keypoint candidates from inspected images are matched to a limited number of potential counterparts (those of the same pattern only). Computational complexity of model keypoint detection is quite high because we have to examine each location using scanning windows in numerous scales covering the whole range (R_1, R_2). Although the moment calculations are reusable, the equations for parameter estimations should be solved separately for each radius. Since model-building operations are usually performed offline, this disadvantage is acceptable. In the next sub-section the issue of online keypoint detection is discussed. This would be important in a real-time search for objects of interest, i.e. in robotic vision applications.
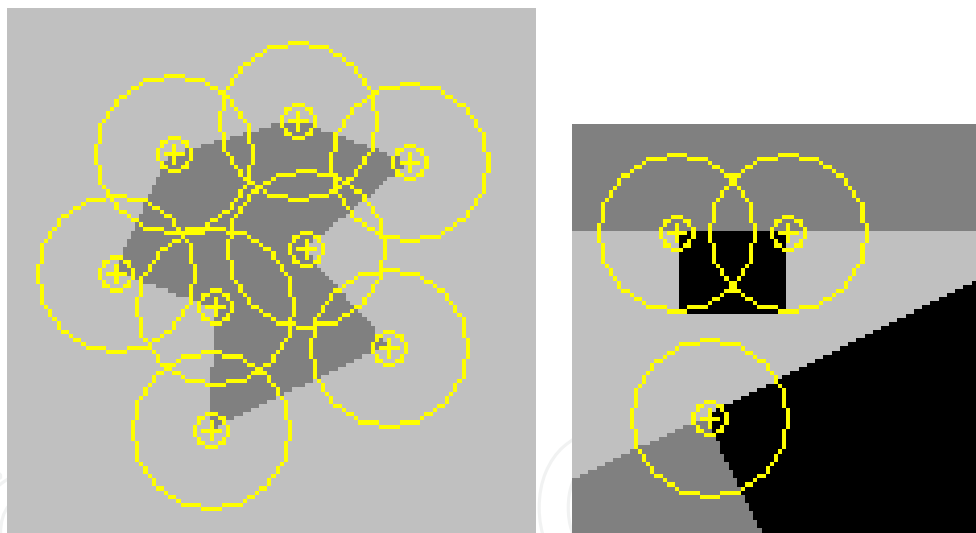

Figure 5. Corner-based model keypoints and 90° T-junction-based model keypoints detected in simple images of good quality. Scanning window radii range from 5 to 20 pixels

Figs 5 and 6 show a few examples of images with model keypoints detected for corner and 90° T-junction patterns. Window radii ranging from 5 to 20 pixels have been used. It should be noticed that in simple images of good quality the model keypoints look prominent to human vision as well. For more complex images, however, many model keypoints look inconspicuously (see Figure 6). Nevertheless, they are also stable features that are consistently present (at least many of them) when the image is distorted.

A   **B**
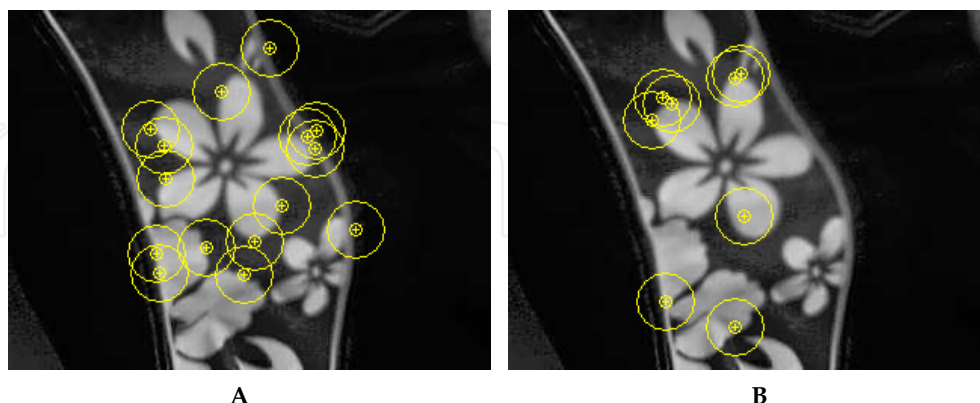
Figure 6. Corner-based model keypoints (A) and 90° T-junction-based model keypoints (B) detected in a more complex image of normal quality. Scanning window radii range from 5 to 20 pixels

2.3 Scene Keypoints for Object Detection

Computational complexity of model keypoints may be too high for real-time applications of machine vision. If, however, similar keypoints can be detected online in inspected images, model keypoints would be very reliable references for matching content of images to the available models. Therefore, we propose a simplified variant of model keypoints, so-called *scene keypoints*. The definition of *scene keypoints* is very similar to Def. 1.

Definition 2

For a given image, pixel (x,y) is an *approximation-based scene keypoint of radius R* (shortly *scene keypoint*) defined by a circular pattern TP if for the scanning windows located at (x,y):

1. The approximations by TP pattern exist for the scanning radius R and for another radius R_{sub}, where R_{sub} is a predefined constant percentage of R (the recommended value for R_{sub} is approx. 70% of R).
2. The approximations parameters obtained for R and R_{sub} radii are similar.
3. If several neighbouring pixels satisfying (1) and (2) exist, the scene keypoint is located at the pixel where the *similarity* between both scanning windows and their approximations reaches a local maximum.

Usually, the practical constraints defined and explained after Def. 1 are also applicable to the above definition.

Computational complexity of detecting scene keypoints is much lower. Moments of only two windows (R and R_{sub} radius) are computed at each location, and reusability of moment calculations both at the current location and for neighbour pixels can be exploited. The equations for parameter identification are also used only twice.

Figs 7 and 8 contain exemplary images with scene keypoints detected (for corners and 90° T-junctions) using windows of radii 10 and 7 pixels. Obviously, for the same images the number is scene keypoints is larger than the number of model keypoints because the detection algorithm is much less restrictive. Even though for perfect-quality images (compare Figure 5 to Figure 7) we would expect the same keypoints, the presence of additional keypoints can be explained by digital effects and mathematical properties of the

moments used. Nevertheless, each model keypoint is also always detected as a scene keypoint

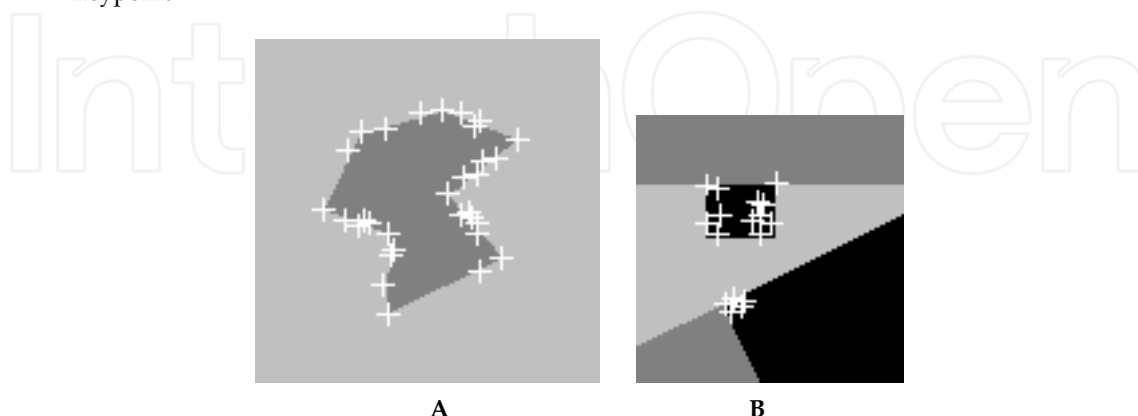**A**   **B**

Figure 7. Corner-based 90° scene keypoints (A) and T-junction-based scene keypoints (B) detected in the images from Figure 5. Scanning window radii are 7 and 10 pixels

A   **B**

Figure 8. Corner-based scene keypoints (A) and 90° T-junction-based scene keypoints (B) detected in the image from Figure 6. Scanning window radii are 7 and 10 pixels

Matching scene keypoints extracted from analysed images to the database model keypoints is the fundamental operation in the proposed object detection framework. The following section discusses practical aspects of matching. In particular, the adaptability of the method (through selection of thresholds and matching rules) is highlighted.

3. Matching Keypoints for Object Detection

Matching keypoints extracted from images to the database keypoints is used in the majority of works where the goal is to identify objects that might be partially occluded or overlapping (e.g. Lowe, 2004; Mikolajczyk & Schmid, 2004; Islam, 2006; etc.). Unfortunately, the numbers of keypoints are usually very large. Typical scenes used for experiments (e.g. Islam, 2006) contain hundreds of keypoints, while the number of keypoints in databases

with just a few objects captured from a reasonable number of viewpoints can easily reach tens of thousands. Thus, the matching procedures become a serious computational problem. In order to optimise the matching and to avoid too many potential matches, researches either propose multidimensional descriptors of the keypoints and/or use carefully designed matching schemes. For example, 128 gradient-based directional descriptors are used in Lowe, 2004, while in Islam, 2006 only five moment descriptors are used but an efficient hashing technique has been developed to speed up keypoint matching.

In the proposed method, the abovemetioned problems are significantly simplified. Even if the overall number of model keypoints is large, they are divided into different categories (defined by different patterns) that can be handled independently. Scene keypoints are similarly divided into the same categories (even though the total number of scene keypoints for typical images may look larger than the numbers seen in other works). Eventually, each scene keypoint is only matched to the model keypoints in the same category which greatly reduced the computational efforts and allows parallelisation of the matching process.

Descriptors of both model and scene keypoints are obviously parameters of the corresponding pattern approximations. Such descriptors can be used more selectively than other descriptors (e.g. Koenderink & van Doorn, 1987; Lowe, 2004; Islam, 2006, etc.) that are based on general properties of image intensities. Generally, the processes of keypoint detection and matching can be adaptively tuned to various applications. Three issues are highlighted below (the problem is scale invariance is separately discussed in Subsection 3.1).

Thresholds

The number of extracted keypoints depends on several threshold values (see Subsections 2.2 and 2.3) defining the acceptable accuracy of pattern approximations and the minimum levels of visual prominence (contrasts) of the scene keypoints. It is possible, for example, to demand high accuracy and to accept very low contrasts. Then the method would be able to identify only those image fragments that are very accurately approximated by the patterns used. However, such fragments may not be even visible to a human eye. Typically, such requirements can be used for search in poorly illuminated scenes (detection of frauds in images may be another application). Alternatively, only highly-contrasted approximations could be accepted as keypoints with less demands regarding the accuracy of the approximations. This would be potentially useful for detecting objects that may be seen differently than in the database images (but the scenes are expected to be well illuminated). Moreover, the level of acceptable differences between the descriptors of model keypoints and scene keypoints determines the overall behaviour of the method (high numbers of keypoints with possibly many false positives *versus* high confidence keypoints only).

Configuration parameters

The configuration parameters of keypoint approximations have a higher priority as they specify geometry of the local structures of the observed scenes. However, the parameters defining rotations of the patterns (e.g. β_2 angle in Figs 1A, 1B and 1D) should be carefully used for matching (unless the search is for objects at certain orientations). Generally, the orientation parameters are used only in later stages (see Section 4). Moreover, parameters indicating distances (e.g. β_1 in Figs 1B and 1C) should be measured both absolutely and relatively to the window radius (for scale invariance, more in Subsection 3.1).
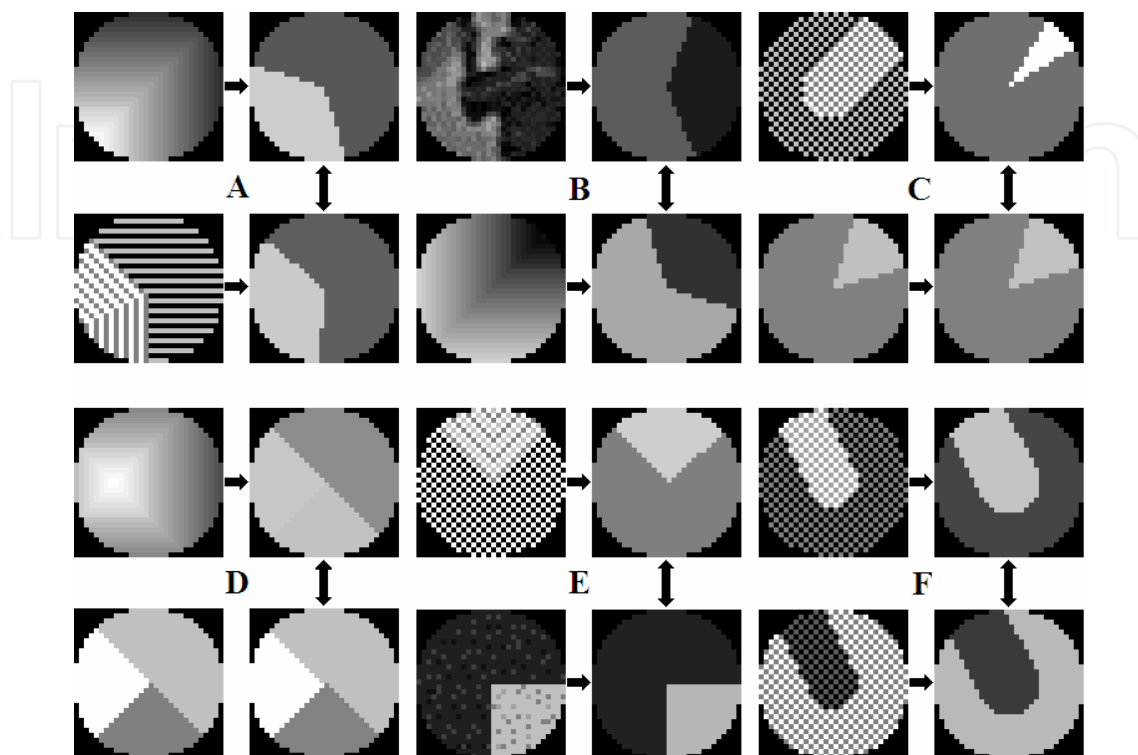
Figure 9. Examples of windows matched using different patterns and for diversified conditions (optimum approximations also shown for references):
(A)   Low accuracy of approximations acceptable. High similarity for angular widths and intensities required. Orientation match ignored.
(B)   Low accuracy of approximations acceptable. High similarity between angular widths required. Only relational match for intensities. Orientation match ignored.
(C)   Low accuracy of approximations acceptable. Similarity between angular widths ignored. Only relational match for intensities. Orientation match required.
(D)   No contrast thresholds in approximations. Low accuracy of approximations acceptable. All configuration parameters matched. Intensity matching not used.
(E)   High accuracy of approximations required. All configuration parameters matched. Intensities matched proportionally.
(F)   Good accuracy of approximations required. High similarity between line widths required. Intensity matching not used

Intensity parameters
The intensity parameters of keypoint approximations can be used more selectively that the configuration parameters (and their significance is usually lower). In the extreme scenarios they are not used in the matching process at all (i.e. only the local structures of the objects are important) although the other extreme is to match them accurately (to detect keypoints viewed in the same illumination conditions). Typically, either only relations between the intensities are verified (e.g. a scene corner keypoint can match a given model corner

keypoint if the acute section is lighter than the obtuse one – see Figure 1A) or the proportions between the intensities of keypoints should match to a certain level.

To illustrate the above issues, Figure 9 presents exemplary pairs of circular windows (they are in the same scale as scale invariance is discussed in Subsection 3.1) that can be matched under various (sometimes not very realistic) assumptions. The windows are already placed at the local maxima of similarity functions so that if keypoints are extracted they would be found at the same locations. The corresponding pattern approximations are also given to highlight that matching is actually performed between the approximations rather than between the original contents of windows.

3.1 Scale Invariance in Keypoint Matching

Although the examples given in Figure 9 address the issue of matching circular windows of the same radius, the same approach can be used for matching *scene keypoints* of the same size. The only difference is that the match should be satisfactorily established both for the outer windows (of radius R) and for their sub-windows (of radius R_{sub}). However, matching objects shown in arbitrary scales to their models (i.e. matching scene keypoints to the model keypoints) can be done only under additional assumptions.

If a "visual correspondence" between a fragment of a model image and a fragment in an inspected image exists, it can be generally confirmed by a match between the corresponding model keypoint (defined for the radius range (R_1, R_2) – see Def. 1) and the scene keypoint (defined by radii R and R_{sub}) only when:

$$\sigma R_1 \leq R_{sub} \quad \text{and} \quad R \leq \sigma R_2 \tag{7}$$

where σ is the *relative scale* between the model image and the processed image.

The relative scale defines how much the size of an object (measured in the image units) has been changed against the size of the same object in the model image. The relative scale is jointly determined by the image resolution, the camera-object distance and the camera focal length. Detailed analysis of relative scale issues in the context of object detection in given in (Saiful, 2006).

In Section 2, we extract exemplary model keypoints using the range of radii (R_1, R_2) from 5 to 20 pixels, while exemplary scene keypoints are found using 10 and 15 pixels. From Eq. (7) we can immediately calculate that for such conditions images of objects of interest can be prospectively matched to the model images if the relative scale changes from 1.3 to 0.33. It means that the objects can be only insignificantly enlarged, but the up to three times reduced in size. These results correspond to requirements of typical applications (e.g. in mobile robotics) where exemplary objects of interest are available so that their images can be captured from a close proximity. In the actual search operations, however, those objects would be usually seen from a longer distance, i.e. the size reduction in captured images is more likely to happen.

Moreover, the approximation parameters representing distances (e.g. β_1 and β_2 in Figure 1C) should be matched is a special way. They are invariant under usage of variable-radius windows in terms of absolute distances, but they are not invariant *relatively* to the radius. Thus, if a scene keypoint is captured in an unknown scale, such parameters cannot be directly matched to the values in model keypoints. However, they can be later used for verifying the validity of the matches (see Sub-section 4.1).

It should be finally remarked that the selection of radius ranges over which the model keypoints are built affects both scale-sensitivity and robustness of object detection. With wider (R_1, R_2) the scale invariance of obviously expanded to more scales. However, the number of model keypoints can be reduced as the pattern approximations must be stable over a wider range of radii. Therefore, the abilities to detect objects (both fully and partially visible ones) deteriorate. For occluded objects, fewer locations corresponding to model keypoints are seen, while for fully visible objects fewer correspondences can be found to verify hypotheses about the presence of objects. Limited (R_1, R_2) results in the opposite effects, i.e. the scale invariance is reduced to a narrower range, but the method is potentially able to detect objects under stronger occlusions and/or in poorer visibility conditions.

4. Framework for Object Detection

4.1 Hypothesis Building and Verification

Generally, keypoint-based object detecting algorithms are voting schemes where an object of interest is considered found if a sufficient number of keypoints are consistently matched to the corresponding model keypoints (e.g. Wolfson & Rigoutsos, 1997). In our method, we propose to use such a method already presented in (Islam, 2006). The method has been applied to different types of keypoints, but it is also naturally applicable (after minor modifications) to the keypoints proposed in this paper.

To detect presence of the objects of interest in processed images, several steps are performed as outlined below. Detailed explanations of the steps are given in (Islam, 2006).

In the first step, clusters of scene keypoints matching the model keypoints are created using Generalised Hough Transform (GHT) similarly to Ulrich et al., 2003. The accumulator of $u \times v$ size is used, where u is the number of objects and v is the number of model images for each object. A scene keypoint falls into an accumulator bin if it matches a model keypoint from the corresponding image. Usually, scene keypoints match several model keypoints (depending on the matching strategy the numbers can be larger or smaller – see Section 3). Each bin that collects a sufficient number of scene keypoints should be considered a hypothesis regarding a presence of the object (seen from a particular viewpoint). All such hypotheses are subsequently verified. It should be noted that eventually several hypotheses can be accepted. If they use different sets of points, such multiple hypotheses indicate the presence of multiple objects in the scene. If two or more accepted hypotheses use similar clusters of scene keypoints and yet produce different results, it means that either partially occluded different objects have similar model keypoints in the visible parts, or different objects accidentally share similar model keypoints.

Simple examples illustrating advantageous and disadvantageous aspects of using such hypotheses are given in Figure 10. The examples are taken from (Islam, 2006) so that different types of keypoints are used, but the same effects can be expected for the proposed keypoints as well.

Hypotheses are verified using the concept of *shape graphs* and *scene graphs*. *Shape graphs* are built for model images while *scene graphs* are built for analysed images; otherwise they are identically defined fully connected graphs. Nodes of the graph for a given cluster of keypoints represent the keypoints (scene keypoints for a scene graph and the matching model keypoints for a shape graph). Each edge of the graph is labelled by the distance between the adjoined nodes (keypoints).

An iterative algorithm is used to find the maximum sub-graphs of a scene-graph and a shape-graph for which all corresponding pairs of edges have approximately proportional label values. This iterative algorithm converges very fast and in most cases only a few iterations are needed. The generated sub-graphs specify the final set of scene and model keypoints used to confirm the validity of the hypothesis. The selected keypoints not only match the model keypoints but also their spatial distributions are similar.
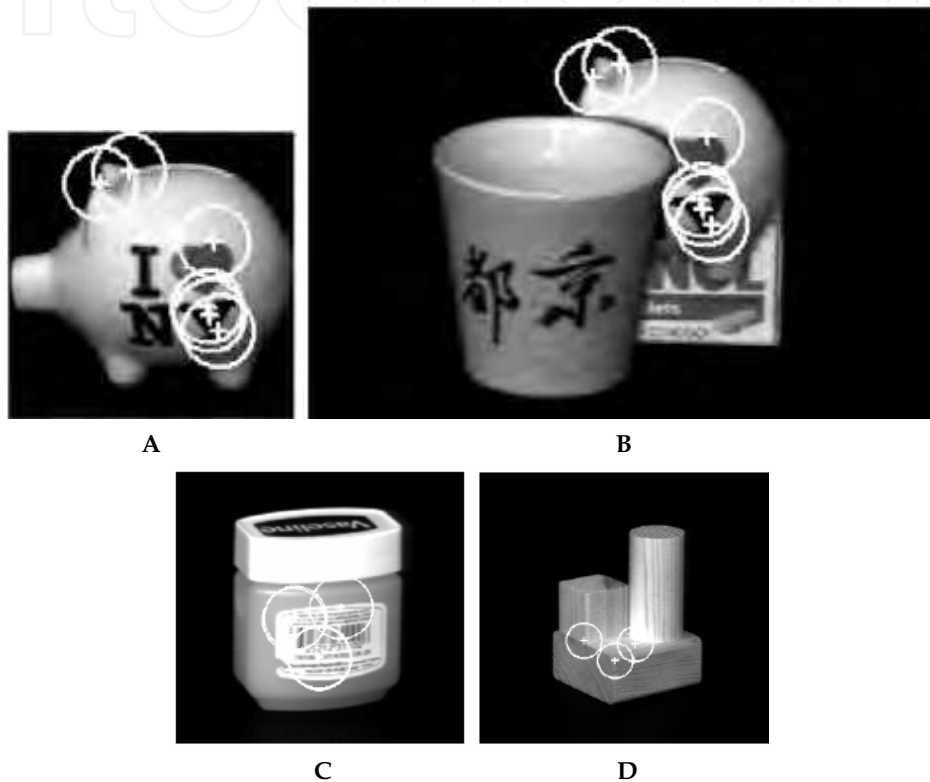

Figure 10. Model images (A and C) successfully matched to test images (B and D, respectively). Clusters of matching keypoints are shown

The minimum number of nodes in the subgraphs (i.e. the number of consistently matched keypoints) required for confirmation of the object's identity may depend on the set of objects under consideration. However, our experiments and statistical analysis show that usually 5 keypoints are enough. It can be noticed, that the incorrect match between Figure 10C and Figure 10D is confirmed only by three keypoints.

The hypotheses verification can be additionally supported by the analysis of configuration parameters of scene keypoints. In particular, only those keypoints from a single cluster would be used for building a scene-graph which are consistently rotated with respect to the corresponding model keypoints (see the last column of Table 2). This is a very powerful constraint that greatly reduces the complexity of the hypothesis verification procedure.

4.2 Exemplary Results

The following example illustrates the process of object detection (i.e. hypothesis verification) briefly explained above. Selected issues regarding keypoint matching are also highlighted.

A   **B**
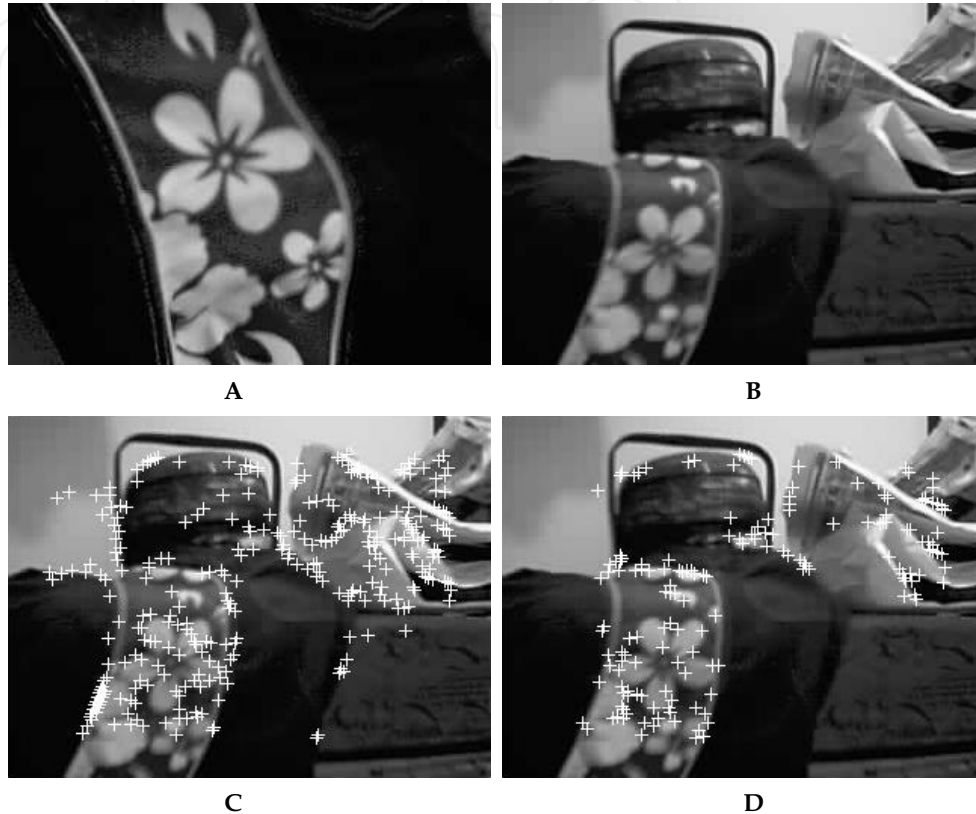
C   **D**

Figure 11. The model image (A) and the test image (B). Corner scene keypoints shown in (C) and 90° T-junction scene keypoints are given in (D)

An exemplary model image and a test image are given in Figure 11. Location of corner scene keypoints and 90° T-junction scene keypoints detected in the test image are also shown.

The selected example deliberately uses a piece of cloth as the object of interest to show that the method has a potential to deal with some non-rigid objects as well. Match results have been obtained using only two types of scene keypoints shown in Figs 11C and 11D. To compensate for non-rigidity of the object, the shape/scene graphs labels have been compared only for the longest edges (so that minor local shape distortions do not affect the hypothesis verification). The additional assumptions are as follows:

- Intensity parameters in scene keypoints and the corresponding model keypoints differ approximately similarly.
- Angular widths in the corner scene keypoints are similar to the angles in the matching model keypoints.
- All scene keypoints should be similarly rotated relatively to the orientations of corresponding model keypoints.

Figure 12. Matched keypoints in the model and test images (⊕ corner keypoints, ■ 90° T-junction keypoints

	A	D	E	I	J	Y
A	×	$\dfrac{40.0}{69.7}=0.57$	$\dfrac{37.6}{68.8}=0.55$	$\dfrac{73.0}{128}=0.57$	$\dfrac{63.0}{104.5}=0.59$	$\dfrac{89.8}{145.3}=0.61$
D	$\dfrac{40.0}{69.7}=0.57$	×	$\dfrac{60.1}{105.6}=0.57$	$\dfrac{72.4}{130}=0.56$	$\dfrac{30.1}{52.5}=0.57$	$\dfrac{51.9}{81}=0.63$
E	$\dfrac{37.6}{68.8}=0.55$	$\dfrac{60.1}{105.6}=0.57$	×	$\dfrac{46.4}{71.9}=0.63$	$\dfrac{69.1}{109}=0.63$	$\dfrac{98.1}{152.4}=0.64$
I	$\dfrac{73.0}{128}=0.57$	$\dfrac{72.4}{130}=0.56$	$\dfrac{46.4}{71.9}=0.63$	×	$\dfrac{57.4}{99.7}=0.57$	$\dfrac{79}{134.9}=0.59$
J	$\dfrac{63.0}{104.5}=0.59$	$\dfrac{30.1}{52.5}=0.57$	$\dfrac{69.1}{109}=0.63$	$\dfrac{57.4}{99.7}=0.57$	×	$\dfrac{31.4}{53.3}=0.59$
Y	$\dfrac{89.8}{145.3}=0.61$	$\dfrac{51.9}{81}=0.63$	$\dfrac{98.1}{152.4}=0.64$	$\dfrac{79}{134.9}=0.59$	$\dfrac{31.4}{53.3}=0.59$	×

Table 1. Distance ratios for the corresponding fragment of the shape graph (denominator values) and the scene graph (numerator values) for Figure 12 images

Figure 12 presents pairs of finally matched keypoints, and Table 1 shows a fragment of the shape/scene graph (only the most distant keypoints are included). Although certain variations of the ratio between the corresponding distances in the shape and scene graphs can be noticed, the average ratio is consistently near **0.6** which can be assumed the approximation of the relative scale between the model image and the test one. This value corresponds to the visual assessment of Figs 11A and 11B.

Corner approximations

	Keypoint type	Intensities	Angular width	Orientation difference
A	model	187 and 57	134°	32°
A	scene	160 and 60	149°	32°
B	model	194 and 46	90°	24°
B	scene	165 and 60	104°	24°
C	model	186 and 45	153°	25°
C	scene	151 and 44	157°	25°
D	model	187 and 48	147°	29°
D	scene	135 and 38	146°	29°
E	model	121 and 18	140°	41°
E	scene	91 and 20	154°	41°
F	model	162 and 35	149°	36°
F	scene	136 and 56	152°	36°
G	model	162 and 38	151°	36°
G	scene	117 and 21	154°	36°
H	model	171 and 53	142°	24°
H	scene	137 and 31	153°	24°
I	model	154 and 10	142°	26°
I	scene	123 and 7	151°	26°
J	model	174 and 48	145°	37°
J	scene	143 and 35	156°	37°
K	model	26 and 172	158°	29°
K	scene	19 and 149	158°	29°
L	model	20 and 169	158°	32°
L	scene	14 and 139	160°	32°

90° T-junction approximations

	Keypoint type	Intensities	Orientation difference
X	model	91, 45 and 179	41°
X	scene	95, 61 and 136	41°
Y	model	142, 4 and 73	32°
Y	scene	138, 9 and 75	32°

Table 2. Approximation parameters for the model and scene keypoints used for the match shown in Figure 12

As a further reference, Table 2 compares parameters of corner approximations and T-junction approximations obtained for model and scene keypoints used for the hypothesis confirmation. It shows a relatively high consistency for the orientation differences (ranging from 24° to 42°) for all keypoints and high level of similarity for the angular widths of corner keypoints. The differences between the corresponding intensities are wider (which is unavoidable for images captured in different conditions) but they are consistent as well. In

particular, if the intensities differ they change in a similar way for all intensities of a given approximation.

5. Concluding Remarks

We have presented principles and exemplary results of a novel technique for detection of known objects in inspected images. The method is based on new types of keypoints which are the focus of this paper. The proposed keypoints are significantly different from typical gradient-based keypoints used in the alternative techniques. Our keypoints are based on moment-derived pattern approximations of circular patches. Though currently only a few patterns are used (i.e. corners, T-junctions and round tips of thick lines) a wide range of other patterns can be added using the approach presented in our previous works (e.g. Sluzek, 2005). The keypoints are characterised by intensity and configuration descriptors (e.g. angular widths and orientation of the approximations) that are generally robust under illumination changes, noise, texturisation, and other typical real-world effects. More importantly, the keypoints are also scale-invariant within a certain range of scales. This has been obtained by using two different methods for keypoint building in model images and in analysed images.

Model images of database (known) objects are processed in multiple scales in order to identify *model* keypoints that are invariantly characterised within the assumed range of scales. The operation may be computationally expensive, but it is typically performed either offline or in the preliminary phase of deployment when timing constraints are not critical. However, the *scene* keypoints extracted from inspected image are based (unlike keypoints used in other scale-invariant techniques) on a single-scale image scanning and processing. Additionally, the efficiency of keypoint matching is improved by a simultaneous usage of several keypoint types. Even if the overall number of keypoints (both model and scene ones) is comparable to the numbers typically extracted and used by other methods, scene keypoints of a certain category are matched only against the corresponding subset of model keypoints of the same category. Therefore, the computational costs of image analysis are relatively low and the method is suitable for real-time applications (e.g. for exploratory robotics which is considered the primary application area).

Several improvements of the method are currently envisaged First, we propose to enhance the efficiency of keypoint matching by adding (without any significant computational costs) more keypoint descriptors. For that purpose, moment-based expressions invariant under similarity transformations and linear intensity changes are considered. Although generally such invariants (proposed for colour images and areas of arbitrary shapes in Mindru et al., 2004) are rather complex, we intend to apply them to circular images only. For circular images, the following expressions have been found invariant under similarity transformations and linear intensity changes. For other shapes of the processed areas they are not invariant, however.

$$\frac{(m_{20} - m_{02})^2 + 4m_{11}^2}{R^2 \left(m_{10}^2 + m_{01}^2\right)} \quad \text{and} \quad \frac{2(m_{20} + m_{02}) - R^2 m_{00}}{R\sqrt{m_{10}^2 + m_{01}^2}} \tag{8}$$

where *R* is the circle radius.

Another prospective continuation of the method is to use colour equivalents of the proposed keypoints (with three colour channels processed separately or jointly). We also consider

hardware accelerators for the moment calculations. Selected moment-computing procedures have been already implemented in FPGA as a feasibility study. The results indicate that with a support of an FPGA accelerator a real-time detection of scene keypoints in a TV video stream is feasible.

The primary area of intended applications for the proposed method is intelligent robotics (exploratory robots in particular). The ultimate goal would be a system that can be shown a physical "known object" and subsequently such objects present in complex cluttered scenes can be detected. However, other areas of applications should be highlighted as well. As some recently published results suggest (e.g. Prasad et al., 2004) image retrieval and/or search in visual databases seems to be a potential application area. Using the proposed keypoints, not only the search for known objects or images can be conducted, but also some image-related frauds can be revealed (e.g. detection of almost invisible highly accurate approximations may indicate image doctoring).

Surveillance and/or security systems are another envisaged area for the developed technique. Since such systems are equipped with more and more embedded intelligence, a system that can identify "known intruders" or "particularly dangerous intruders" is a possible scenario. Development of a sensor network with vision capabilities that can eventually incorporate the proposed method has been reported in our recent papers (e.g. Sluzek et al., 2005).

6. References

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, Vol.94, No.2 (Apr. 1987), pp. 115-147, ISSN: 0033-295X.

Edelman, S. (1997). Computational theories of object recognition. *Trends in Cognitive Sciences*, Vol.1, No.8 (Nov. 1997), pp. 298-309, ISSN: 1364-6613.

Harris, C. & Stephens, M. (1988). A combined corner and edge detector. *Proceedings of 4th Alvey Vision Conference*, pp. 147-151, Manchester, Sep. 1988.

Häusler, G. & Ritter, D. (1999). Feature-based object recognition and localization in 3D-space, using a single video image. *Computer Vision & Iage Understanding*, Vol.73, No.1 (Jan. 1999), pp 64-81, ISSN: 1077-3142.

Hueckel, M.H. (1973). A local visual operator which recognizes edges and lines, *Journal of ACM*, Vol.20, No.2 (Apr. 1973), pp 350, ISSN: 0004-5411.

Huttenlocher, D.P. & Ullman, S. (1990). Recognizing solid objects by alignment with an image. *Int. Journal of Computer Vision*, Vol.5, No.2 (Nov. 1990), pp 195-212, ISSN: 0920-5691.

Islam, M.S.; Sluzek, A. & Zhu, L. (2005). Detecting and matching interest points in relative scale. *Machine Graphics & Vision*, Vol.14, No. 3 (Nov. 2005), pp. 259-283, ISSN: 1230-0535.

Islam, M.S. (2006). Recognition and localization of objects in relative scale for robotic applications. *PhD Thesis,* School of Comp. Engineering, Nanyang Technological University (Dec. 2006), Singapore.

Koenderink, J.J. & van Doorn, A.J. (1987). Representation of local geometry in the visual system. *Biological Cybernetics*, Vol.55, No.6 (March 1987), pp. 367−375, ISSN: 0340-1200.

Lowe, D. (2004). Distinctive image features from scale-invariant keypoints. *Int. Journal of Computer Vision*, Vol.60, No.2 (Nov. 2004) pp. 91-110, ISSN: 0920-5691.

Mikolajczyk, K. & Schmid, C. (2004). Scale & affine invariant interest point detectors. *Int. Journal of Computer Vision*, Vol.60, No.1 (Oct. 2004) pp. 63-86, ISSN: 0920-5691.

Mindru, F.; Tuytelaars, T.; van Gool, L. & Moons, Th. (2004). Moment invariants for recognition under changing viewpoint and illumination. *Computer Vision & Image Understanding*, Vol.94, No.1-3 (April 2004) pp 3-27, ISSN: 1077-3142.

Moravec, H. (1983). Stanford cart and the CMU rover. *Proceeding of the IEEE*, Vol.71, No.7 (July 1983), pp. 872-884, ISSN: 0018-9219.

Prasad, B.G.; Biswas, K.K. & Gupta, S.K. (2004). Region-based image retrieval using integrated color, shape, and location index. *Computer Vision & Image Understanding*, Vol.94, No.1-3 (April 2004) pp 193-233, ISSN: 1077-3142.

Rosin, P.L. (1997). Measuring corner properties. *Computer Vision & Image Understanding*, Vol.73, No.2 (Feb. 1999), pp 291-307, ISSN: 1077-3142.

Schmid, C. & Mohr, R. (1995). Local grayvalue invariants for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.19, No.5 (May 1997), pp. 530-535, ISSN: 0162-8828.

Sluzek, A. (2005). On moment-based local operators for detecting image patterns. *Image and Vision Computing*, Vol.23, No.3 (March 2005), pp 287-298, ISSN: 0262-8856.

Sluzek, A; Palaniappan, A. & Islam, M.S. (2005). A wireless sensor network for visual detection and classification of intrusions, *WSEAS Transactions on Circuits and Systems*, Vol.4, No.12 (Dec. 2005), pp 1855-1860, ISSN: 1109-2734.

Sluzek, A. (2006). An improved detection algorithm for local features in gray-level images. In: *Computer Vision and Graphics* (*Computational Imaging and Vision*, vol.32), K.Wojciechowski et al. (Eds.), pp 406-412, Springer, ISBN: 1-4020-4178-0, Dordrecht.

Tarr, M.J.; Bülthoff, H.H.; Zabinski, M. & Blanz, V. (1997). To what extent do unique parts influence recognition across changes in viewpoint? *Psychological Science*, Vol.8, No.4 (July 1997), pp. 282-289, ISSN: 0956-7976.

Ulrich, M.; Steger, C. & Baumgartner, A. (2003). Real-time object recognition using a modified generalized Hough transform. *Pattern Recognition*, Vol.36, No.11 (Nov. 2003), pp. 2557-2570, ISSN: 0031-3203.

Wolfson, H.J. & Rigoutsos, I. (1997). Geometric hashing: an overview. *IEEE Computational Science & Engineering*, Vol.4, No.4 (Oct. 1997), pp. 10-21, ISSN: 1070-9924.

Vision Systems: Applications

Edited by Goro Obinata and Ashish Dutta

Computer Vision is the most important key in developing autonomous navigation systems for interaction with the environment. It also leads us to marvel at the functioning of our own vision system. In this book we have collected the latest applications of vision research from around the world. It contains both the conventional research areas like mobile robot navigation and map building, and more recent applications such as, micro vision, etc.The fist seven chapters contain the newer applications of vision like micro vision, grasping using vision, behavior based perception, inspection of railways and humanitarian demining. The later chapters deal with applications of vision in mobile robot navigation, camera calibration, object detection in vision search, map building, etc.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Andrzej Sluzek and Saiful Islam (2007). New Types of Keypoints for Detecting Known Objects in Visual Search Tasks, Vision Systems: Applications, Goro Obinata and Ashish Dutta (Ed.), ISBN: 978-3-902613-01-1, InTech, Available from:
http://www.intechopen.com/books/vision_systems_applications/new_types_of_keypoints_for_detecting_known _objects_in_visual_search_tasks

INTECH

open science | open minds