Proceedings of the 6th WSEAS International Conference on Robotics, Control and Manufacturing Technology, Hangzhou, China, April 16-18, 2006 (pp187-192)

On Fusion of Gated Imaging and Local Shape Features for Identification of Objects in Cluttered Scenes

ANDRZEJ SLUZEK[1,3], MD SAIFUL ISLAM[1], TAN CHING SEONG[2]
[1]School of Computer Engineering,
Nanyang Technological University, Blk N4 Nanyang Avenue, SINGAPORE
[2]Singapore Institute of Manufacturing Technology, 71 Nanyang Drive, SINGAPORE
[3]SWPS, ul.Chodakowska 19/31, Warszawa, POLAND

Abstract: - We discuss a novel method of visual detection of rigid objects (that are known to the system) randomly located in complex cluttered scenes (possibly in a turbid medium that additionally deteriorates visibility). The images are captured using a gated imaging system so that (1) the medium backscattering noise is minimized and (2) only the scenes components within a predefined distance from the camera are captured in the images while the rest of the scene (containing presumably objects that are not of immediate interest) remains invisible. The database of known objects is built (in *reference scale*) using local shape features (interest points) extracted from template images showing the objects of interest from various viewpoints. By matching interest point detected in gates images (*relative scale* is used there) to the local features from the database, known objects can be identified (even if only partially visible). The paper briefly discusses the proposed methodology and explains why complexity of vision-based navigation algorithms could be dramatically reduced (while the robustness is equally dramatically improved) by adopting the proposed approach.

Key-Words: - object identification, gated imaging, moment invariants, local shape, relative scale

1 Introduction

In many applications of intelligent robotics, there is often a requirement to visually identify objects in very complex, unpredictable terrains and possibly in reduced-visibility conditions (e.g. turbid water, dense smoke, fog, etc.). Typical examples are identification of landmarks for autonomous navigation or distinguishing between obstacles and objects of interest. Generally, such objects are present in cluttered environments (so that partial occlusions are possible) and can be visually distorted (compared to their models) by various geometric and photometric transformations. The identification process should be robust to all such conditions and transformations.

In this paper we discuss a methodology that exploits advantages of two recently developed techniques: i.e. gated imaging [1] and local shape features defined in *reference scale* [2] are combined into a scheme that has a great potential of satisfying the above requirements.

The purpose of gated imaging is to reduce the amount of visually acquired data by range discrimination (only objects within a certain distance from a camera – see Section 2). The *relative scale* is an alternative to the multi-scale approach. When combined with gated imaging, the relative scale allows very efficient processing and matching for local shape features regardless the size of objects in the image (details are given in Section 3). Section 4 of the paper discusses advantages and possible limitations of the combination of both methods (with an example illustrating the expected difficulties). Section 5 concludes the paper and highlights the future directions as only the preliminary results have been achieved so far.

2 Basics of Gated Imaging

Range-gated imaging is a relatively novel method of image acquisition with devices that can discriminate between reflected and scattered visual signals (e.g. [1], [3], [4]). A gated imaging system typically consists of a pulsed laser (with the pulses usually diverged into a conical shape) a high-speed gated camera, and the control and synchronization circuitry. Projected pulses reflect from objects and return to a camera with electronically controlled (gated) shutter. If the gate opening is synchronized with the head of the pulse returning to the camera after reflecting from an object, and closing is synchronized with the pulse tail, the camera captures the image of this object.

Reflections from more distant objects do not return before the shutter closes (i.e. such objects are invisible) while reflections from front objects return before the shutter opens (i.e. such objects become black shadows). The principles of operation are explained in Fig.1, and an exemplary non-gated image is compared to two gated images (captured at different ranges) in Fig.2.

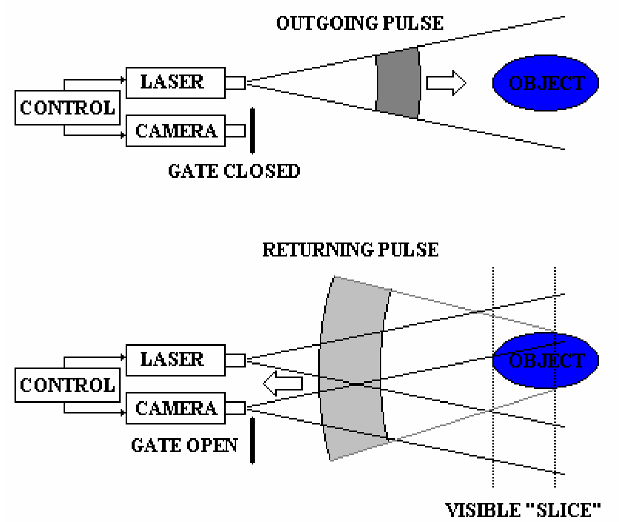

Fig.1. Principles of operation of gated imaging.

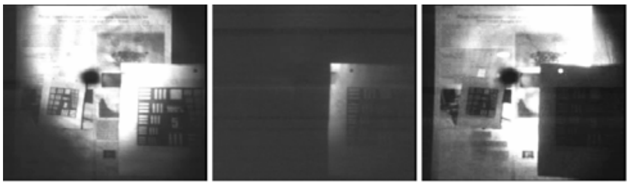

Fig.2. A non-gated image compared to two gated images.

Additionally, in turbid conditions gated imaging minimizes the backscattering noise as most of it (produced by the medium between the camera and the object) returns when the shutter is closed. Fig.3 shows typical time-domain profiles of the signal intensity returning to a camera after illuminating the scene by a short laser pulse in water of various turbidities. It clearly illustratess that when the camera shutter is open only during the rightmost peak (object reflection) the backscattered signal (the leftmost peak) is minimized.

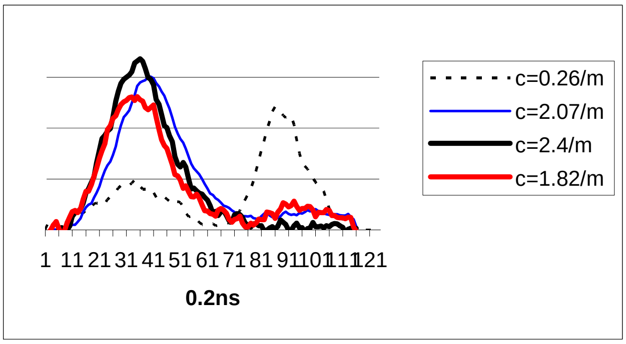

Fig.3. Returning signal profiles for a 1.8m opaque target illuminated by a 9ns laser pulse in water of various turbidities (attenuation *c*).

From the autonomous navigation perspective, the most attractive property of gated imaging is its ability to be "tuned" to a certain range so that no matter how complex the scenes is, the objects become visible only if they are within that range.

3 Local-shape Features in a Relative Scale

Local-shape features are a generally accepted tool for object detection in cluttered scenes (e.g. [5], [6]). This is because no segmentation of objects is required and the approach is robust to partial occlusions and background changes.

Other issues to deal with in actual applications are illumination changes and pose changes (geometric transformations of the object in the image). Particularly, the change of scale of the object is a concerning issue for the local-shape methods. In order to cope with the scale changes, multi-scale methods (pyramids or scale-space) have been proposed. Unfortunately, the multi-scale methods require huge memory for object modeling, and the object recognition process is computationally expensive.

3.1 Interest point detection ad description

In majority of applications, local-shape methods are based on detection of Harris interest points, e.g. [7]. Interest points are easily perceivable discontinuities of image intensity (usually associated with corners or corner-like features). Fig.4 shows a simple image with interest points detected automatically by the Harris detector.

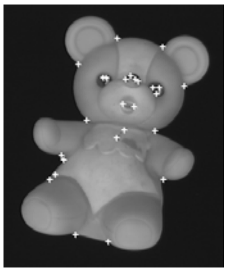

Fig.4. Interest points extracted by the Harris detector.

We propose the *reference/relative* scale approach where interest points are extracted using the scale normalized Harris detector

$$N(X, \sigma_I)=\sigma_D^2\, g(\sigma_I)\otimes\begin{bmatrix} I_x^2(X,\sigma_D) & I_xI_y(X,\sigma_D) \\ I_xI_y(X,\sigma_D) & I_y^2(X,\sigma_D) \end{bmatrix} \qquad (1)$$

where X is a point of an image, and I is the image intensity function. The scale is represented by σ_I , i.e. $g(\sigma_I)$ is a circular Gaussian integration window with σ_D variance (proportional to σ_I). The interest points are defined as points where the value of

$$C(X, \sigma_I) = \det(N(X, \sigma_I))-\lambda \mathrm{tr}^2(N(X, \sigma_I)) \qquad (2)$$

exceeds a predefined threshold.
The number of interest points detected in a given view of an object depends on viewing aspects (scale, orientation) and illumination conditions. To solve this problem, we have proposed a method (details is [2]) that assures a high rate of repeatability in interest point detection under changes of scale and illumination variations. The method's performance is superior to results reported by others (e.g. [7]).
To characterize interest points (and to match them to the corresponding interest points from other images) we have proposed a simple yet very powerful local descriptor (LID) that is invariant under 2D similarity transformations (rotation, scaling) and contract changes. LID is computed within a circular neighbourhood of each interest point detected.
The invariant properties of LID are obtained by using moment expressions invariant to geometric and photometric transformations. Hu moment invariants (see [8]) are the starting point that provides geometric invariance. The first Hu invariants are:

$$\varphi_1 = \eta_{20} + \eta_{02} \tag{3}$$

$$\varphi_2 = (\eta_{20} - \eta_{02})^2 + 4(\eta_{11})^2 \tag{4}$$

$$\varphi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \tag{5}$$

from which so-called Maitra invariants (see [9]) can be derived. Maitra invariants are additionally invariant to photometric changes (contrast variations). The first Maitra invariants are:

$$\beta_1 = \frac{\sqrt{\varphi_2}}{\varphi_1} \tag{6}$$

$$\beta_2 = \frac{\varphi_3}{\varphi_2} . \frac{\mu_{00}}{\varphi_1} \tag{7}$$

Our experiments have indicated that the simplest Maitra invariant β_1 has the best information content (see [2]) combined with the lowest noise sensitivity so that is has been selected for building LID.
Actually, LID is computed for *n* concentric circular patches around the interest point so that a higher dimensionality (i.e. a better discriminatory quality) can be achieved. Fig.5 illustrates the idea. We typically use 3 to 5 circles.

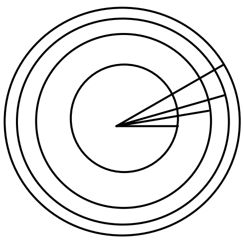

Fig.5. Concentric circles centered at the interest point. The radii of the circles increase slowly.

3.2 Modeling and detection of 3D objects

Typical objects of interest are 3D solids. Therefore, they cannot be modeled using a single image. Such objects are modeled using interest points computed from multiple views of the object. Several images of the object are captured from significantly different viewpoints, as a representative set of *reference images* for the whole object. The number of required views depends on the object itself. We only consider views from a constant distance with uniform intervals (e.g. 15°) of orientation angles. Arbitrary reference scale is assigned to all reference images. The proposed LID for each of detected interest points is subsequently computed and memorized in the database of modeled objects.
Additionally, for each reference image we compute spatial relations between interest points. These spatial relations are represented by a connected graph called *shape-graph*, which allows partial matching. If the viewpoint changes are insignificantly small, the distances between interest points are proportional to the scale of the object, and remain invariant to other transformations of the image.
When an object of interest is subsequently present in a camera-captured image, we expect that the interest points extracted from the image can be matched to the database interest points. However, even for simple objects the overall number of interest points in multiple reference images is very large (hundreds or thousands). Fig.6 shows an exemplary reference image and an exemplary camera-captured image with interest points detected.

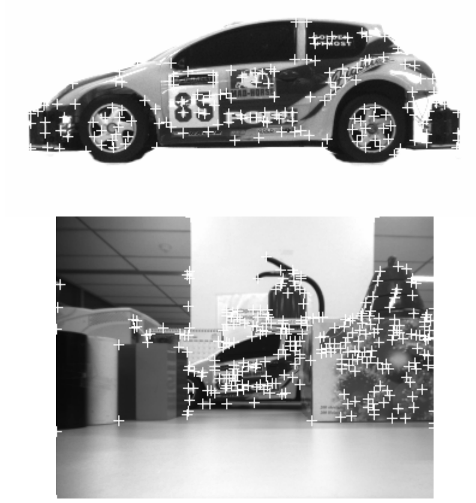

Fig.6. Interest points detected in a reference image, and in an exemplary test image.

Thus, the matching process for interest points extracted from camera-captured scenes could be prohibitively time-consuming (in unstructured terrains the majority of interest points belong to the background/foreground rather than to the objects).

To solve this problem, we have implemented an efficient hashing technique. Our hashing technique, called *PCA-hashing***,** uses principal component analysis method to obtain the most uniform distribution of data. During the off-line process of object modeling we accumulate all the LIDs extracted form the reference images and select proper hash-key to create a lookup table. Then, multidimensional LIDs are properly indexed to optimize the search during interest point matching.
Now, matching between interest points extracted from a camera image and the database interest points can be performed very efficiently, *if* the relative scale is known. This is an important assumption since LIDs are computed within circles proportional to the relative scale. If an incorrect scale is used the interest point LIDs would be computed over windows with different contents and the match would fail (Fig.7 clearly explains this effect).
In the work presented in [2], the relative scale of the scene objects is either known, or we use an estimate so that only a few scales have to be used in order to finally match the interest point in the correct (at least approximately) scale.

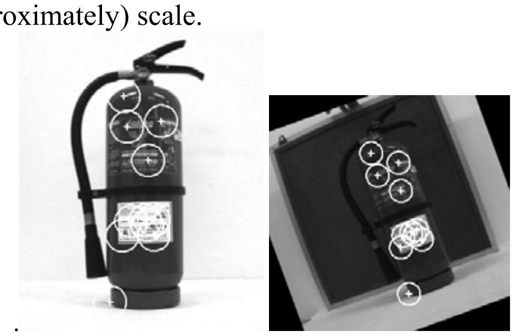

Fig.7. Interest points matching results. The size of LID windows (corresponding to the object scale) is shown.

When a camera image is tested for containing the objects of interest, the image interest points would match the corresponding points of the reference images (if the correct relative scale is used) but there could be also numerous false matches so that a large number of hypotheses regarding the presence of objects can be expected. The hypotheses about the target object and its orientation are generated using generalized Hough transformation (GHT), e.g. [10]. GHT is an extension of the classical Hough transformation used to detect arbitrary shapes.
To verify each of the proposed hypotheses, we check if the *shape-graph* spatial relations between the database interest points are correspondingly satisfied for the image interest points. Only a small number of correctly matched geometric relations (so-called *seed*) is sufficient to ensure the identity of the object in the scene, i.e to verify the hypothesis.
Exemplary final matching results (corresponding to Fig.6 scenario) are given in Fig.8. Note that the relative scale is significantly different than the reference scale.
The computational complexity of the relative scale method is much lover than for alternative multi-scale methods, although the scale should be known (which is not necessary for multi-scale methods). Therefore, the method can be prospectively applied in vision-guided navigation systems. More details are discussed in [2].

Fig.8. Matched pairs of interest points (after hypothesis verification) for Fig.5 scene.

4 Object Detection in Gated Images

In spite of a successful implementation of the relative scale method for object identification and localization in complex scenes, the following weak points of this approach should be highlighted:

a) Not always the relative scale can be accurately estimated so that a large number of various scales could be potentially tested.
b) The number of irrelevant interest points in complex scenes is usually very large (exceeding the number of such points within the objects) so that numerous false hypotheses would be generated and verified.
c) With the increasing number of known objects the analysis of camera-captured images becomes prohibitively time-consuming and/or complex.

It is highly recommended, therefore, to provide: (a) reliable methods for determining the actual relative scale and (b) methods for detecting interest points only within the relevant sections of images. If both goals can be achieved, the overall computational complexity would be significantly reduced so that (c) the number of known objects can be increased without saturating the system.

We propose to use gated imaging as the solution to both (a) and (b) problems. The relative scale of objects clearly depends on the distance to the camera capturing images. With gated imaging, the only visible scene components are those within the "slice of visibility" (see Fig.1). Our experiments have shown that 10% variations in the estimate of relative scale do not affect the performances of interest point matching. Therefore, width of the "visibility slice" should be set according to this requirement.

For example, we assume that a vision-based navigation system (e.g. for a mobile robot) should react to the presence of objects 2m away (a critical distance for collision avoidance). In such a scenario, the "visibility slice" should be 20cm wide, i.e. in the air the laser pulse length or the shutter opening time (or both, which is the most recommended solution) would be 1.3ns (a feasible value for modern gated imaging systems). The shutter opening would be delayed 13ns with respect to the head of the illumination pulse. If the same situation is happening in an underwater environment (with lower speed of light) the duration of the pulse and/or shutter opening can be 1.8ns, and the shutter opening would be delayed by 18ns.

If the distance of object detection and recognition is not fixed, the gating parameters for image capturing should be adaptively changed. Usually, it is difficult to change the pulse duration (in typical gated imaging system the laser parameters are fixed) but both the shutter opening delay and duration can be controlled (within certain limits defined by the camera specification). Digital signals are recommended for controlling the shutter. We have proposed and implemented an FPGA-based system for a direct control of the camera (i.e. width and delay of the shutter opening). Although the system accuracy was limited by the maximum clock frequency available in the FPGA (80MHz in Xilinx Spartan-II with 200,000 equivalent gates) but at the same time the validity of this concept has been strongly confirmed. We have demonstrated that FPGA can control the most important functions of the system, i.e. timing and synchronization of gating. In addition, FPGA can perform various image processing tasks, both in the image domain (enhancement of individual gated images) and the time domain (e.g. fusion of several gated images, ambient light removal, etc.). The results show that FPGA-based control is a feasible solution for gated imaging systems working in real-time applications.The details are available in [11].

If the "visibility slice" is set to the requested distance and width, the only interest point extracted in the camera-captured images would be located within the prospective objects of interest that should be identified (if their models exist in the database, i.e. if they are known to the system). In our experimental scenes, usually not more than 10-20% of interest points have belonged to the objects of interest present in the scenes. Thus, the number of match hypotheses would be greatly reduced with the corresponding increase of the system performances.

There is, however, a certain concern about the matching performances for interest points detected in gated imaging. LID used to match interest points (Section 3.1) is invariant to geometric transformations and linear contrast fluctuations. In gated images, there is often a saturation effect, especially for highly reflective objects illuminated from a close distance by a laser pulse. Two circular windows in Fig.9 illustrate this effect.

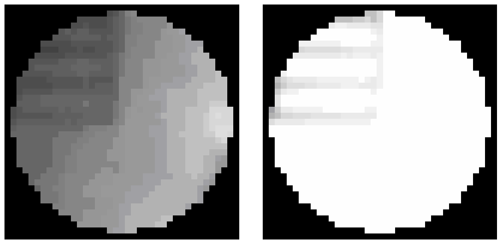

Fig.9. Non-linear photometric transformation between intensities if two corresponding interest point neighbourhoods.

Then, the photometric mapping between intensities of the reference image and the camera image is not linear. If the amount of non-linearity is sufficiently significant, LIDs created using the Maitra invariant (see Eq.6) will not be similar to LIDs of the corresponding interest points.

The hardware-based solution to this problem is to limit aperture of the camera so that the captured gated images are always underexposed and saturation effects never (or almost never) happen. This may reduce, however, the visibility of darker objects of interest even if they are present within the "visibility slice". Another alternative is to develop a new type of local-shape descriptors that would remain invariant (or at least would produce similar values) for circular windows that are partially overexposed (or underexposed).

5 Conclusions

In this paper we have overviewed the concept of combining gated imaging techniques with local shape features (interest points) for detection and identification of known rigid objects in complex environments where partial occlusions and numerous background/foreground objects are probable.

The original relative scale method is used for local modeling (with interest points and their geometrical relations) 3D rigid objects and for subsequent identification of such objects in observed scenes. It is accomplished by detecting a certain number of matching interest points in the captured image, and it works well for images captured by ordinary cameras. However, the main limitation is a large number of irrelevant interest points in the image backgrounds so that numerous matches are detected and all corresponding hypotheses about the presence of objects should be verified. Moreover, in some cases the actual relative scale of the image objects is not known so that the search for matches must be done within a certain range of scales. Both factors slow down the analysis of captured images, sometimes beyond the time needed by the application (which typically would be a vision-guided autonomous navigation system for mobile robotics).

We propose to use gated imaging with controllable gating delay and shutter opening time so that only objects within precisely defined ranges are seen. This would reduce the number of interest point to process (and possibly most of them would belong to the objects of potential interest) and additionally the relative scale is very accurately specified. Altogether, the computational costs of image analysis could be tremendously reduced.

The paper contains only the preliminary study of the proposed approach, explains its advantages, and indicates possible inconveniences and limitations.

A similar concept (though using a different method of extracting objects from the background) has been proposed in other papers (more related to sensor networks) – see [12]. That method, however, is not able to determine the actual relative scale of observed objects, although some indirect estimates can be deducted.

In the future we will focus on further development of gated imaging methods for recognition of known objects. With possible hardware accelerators (already developed for selected tasks in gated imaging) it seems to be a very attractive alternative for high-speed autonomous robotics guided by vision sensors.

Acknowledgements

The authors gratefully acknowledge the support and funding from Intelligent Systems Centre and from Robotics Research Center, Nanyang Technological University.

References:

[1] Tan, C.S., The backscattering phenomena on range-gated imaging in short-range turbid media. PhD thesis, NTU, Singapore, 2005.

[2] Islam, M.S., Relative scale method to recognize and locate objects for robotic applications. *PhD thesis*, NTU, Singapore, 2005.

[3] Fournier, G.R., Bonnier, D., Forand, J.L., Pace, P.W., Range gated underwater laser imaging system. *Optical Engineering*, Vol. 32, 1993, pp 2185-2189.

[4] Swartz, B.A., Laser range gated underwater imaging advances. *Proc. IEEE Oceans'*94, Brest, Vol. 2, 1994, pp. 722–727.

[5] Hausler, G., Ritter, D., Feature-based object recognition and localization in 3D-space using a single video image. *Computer Vision and Image Understanding*, Vol. 73(1), 1999, pp 64-81.

[6] Jurie, F., Schmid, C., Scale-invariant shape features for recognition of object categories. *Proc. IEEE Conf. on Computer Vision and Pattern Rec.*, Vol. 2, 2004, pp II.90–II.96.

[7] Schmid, C., Mohr, R., Bauckhage, C., Evaluation of interest point detectors. *Int. Journal of Computer Vision*, Vol. 37(2), 2000, pp 151-172.

[8] Hu, M.K., Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory*, Vol.8, 1962, pp 179-187.

[9] Maitra, S., Moment invariants. *Proceedings of IEEE* , Vol. 67(4), 1979, pp 697-699.

[10] Ballard, D.H., Generalizing the Hough transform to detect arbitrary shapes. *Pattern Recognition*, Vol. 13(2), 1981, pp 111-122.

[11] Sluzek, A., Fujishima, H., Tan, C.S., Real-time digital control in gated imaging", *Proc. 5th EURASIP Conference. on Speech and Image Proc.*, Smolenice, 2005, pp. 201-206.

[12] Sluzek, A., Palaniappan, A., Islam, M.S., A wireless sensor network for visual detection and classification of intrusions, *WSEAS Transactions on Circuits and Systems*, Vol. 12(4), 2005, pp 1855-1860