Landmark Detection for Vision-based Navigation using Multi-Scale Image Techniques

F. Ferreira, University of Aveiro, Portugal, cfferreira@mec.ua.pt
V. Santos, University of Aveiro, Portugal, vsantos@mec.ua.pt
J. Dias, ISR, University of Coimbra, Portugal, jorge@isr.uc.pt

. . .

ABSTRACT

This article concerns with visual based guidance of a mobile robot platform. A multi-scale image analysis followed by a filtering procedure using disparity information is presented for the detection of landmarks. Good preliminary results at the prototype stage demonstrate the usefulness of a feature extraction method for object recognition as a component of a robot navigation system. An extension of the procedure to applications with stereo is developed. This affords an evaluation of the object regions that generate the extracted features and the elimination of feature points that arise from regions of partial object occlusion.

KEYWORDS: Linear scale-space, stereo, disparity gradient.

I. INTRODUCTION

Navigation is generally defined as controlling motion to arrive at a known location. Recently methods based in form of appearance based method have been adapted for navigation using vision systems, for object recognition and for interpretation of deformable objects such as faces.

The development of computational solutions for simulation of visual behaviors deals with the problem of integration and cooperation between different computational processes for control of active vision systems. In general these computational processes could emulate different and specific artificial visual behaviors and their integration presents distinct facets and problems. First is the interaction and cooperation between different control systems and the second is the use of a common feedback information provided by images captured.

The difficulty of vision problems related to image interpretation, matching and recognition include the ability to find the same interest points under different viewpoint and illumination conditions. Stated a little more formally the task is to identify points of interest that contribute to the creation of repeatable measures for as wide a range of camera transformations as possible. This would include Similarity and, to some extent, Affine transformations and image amplifications.

Whatever the application, we are drawn to the question of what constitutes geometric invariants in an image. In Schmid and Mohr [1], it is mentioned that the only image invariants are the curvature of the isophote line and the flow line. This same article goes on to suggest that this property is of little practical value given that the associated calculations are difficult, and that the noise in real images and their limited resolution negates the remaining usefulness of the result. So a lot of research goes into the extraction of 'partially invariant' features in images. Invariance to scale variations is achieved differently.

Our interest in multi-scale image analysis originates from a robot mission programming approach being developed at our mobile robot laboratory. In this article the essentials of linear scale-space theory will be presented in section II followed by section III which introduces image pyramids and some ways of describing 'points of interest' in the image. In section IV our approach to the extraction of features from a set of images obtained by consistent scale-space treatment followed by section V which describes initial attempts to incorporate range information in the scale-space treatment of images. The description of the experiments conducted, the results

obtained and finally a few conclusions and pointers for ongoing and future work are presented in sections VI and VII.

II. BASIS OF LINEAR SCALE SPACE

The foundations of linear Scale-Space theory are laid in two papers by Koenderink [2] and Witkin [3]. Important later developments came from Lindeberg [4], Romeny [5] among others.

The goal of a linear, multi-scale image analysis method is, in the absence of information other than that contained in an image, to progressively blur out fine details so that other, earlier subtle, image features begin to dominate. This is done by embedding the image in a family of functions related through the variation of a single parameter. Equation (1) is the general 2-d case governing the scale-space treatment. The diffusion equation is found to hold also for polynomial functions of the derivatives, giving us equation (2) where $D_{x^i y^j}(.,.,t) = \frac{\partial L(.,.,t)}{\partial x^i y^j}$.

$$\frac{\partial L}{\partial t} = \frac{1}{2}\partial \nabla^2 L \tag{1}$$

$$\frac{\partial D_{x^i y}(.,.,t)}{\partial t} = \frac{1}{2}\nabla^2 D_{x^i y}(.,.,t) \tag{2}$$

In an extension to the 'Local N-jet', polynomial derivative functions, homogeneous in order of derivatives are used to identify image characteristics. The conversion from the image coordinate system to local gauge coordinates allows flexibility in the comparison of the derivatives and the N-Jet at various points and scales. The approach consists of 1) Normalization of Derivatives across scale and 2) an automatic Scale-selection mechanism that select the scale at which the features assume maximum values along the scale dimension.

Articles by Lindeberg such as [6], [7], [8] deal with the development of an extension of the scale-space theory for the discrete image case. A function containing the modified Bessel functions of integer order is developed as the 'discrete analog of the Gaussian'. This function is defined by the parameter t in equation (3), with t as a proxy for increased blurring or coarser scale. The diffusion equation also leads to the relation that the detection scale for the same object in a zoomed image f' is related to detection scale in the original image f through equation (4), s being the zoom applied.

$$T(n,t) = e^{-t} * I_n(t) \tag{3}$$

$$f(x,t) = f'(sx, s^2 t) \tag{4}$$

III. MULTI-SCALE IMAGE ANALYSIS

Image pyramids have been utilized for quite some time in order to take advantage of memory savings and faster processing as information redundancy in an image increases. Lowe [9] utilizes a sampled Gaussian kernel to smooth an image repeatedly. Each smoothing by convoluting with a Gaussian filter is followed by the reduction of the resolution of the image. The Difference of Gaussian (DoG) image as an approximation of the Laplacian is used to identify the interesting points. The SIFT (Scale-Invariant Feature Transforms) method then creates features that attempt to be invariant to changes in the illumination, viewpoint, translations and partial image occlusions. Feature extraction is followed by the matching using a K-d tree to search and match vectors if length 128, a confirmation being obtained using a variant of the Hough transform followed by a procedure outlined in [10] for transformation parameter estimation.

Lowe [11] describes algorithms for data organization that push the object recognition capabilities (3-D rotation) of the method. These include collecting multiple views of the point and tracking observable features in time. Through the tracking of multiple views of the same object

(variations in angle of view of more that 20 degrees) larger invariance stability of object and scene detection is obtained. Decision theory and conditional probability are also introduced to extend the range of applications [10], [12], [13].

Schmid constructs a 'Local-jet' and extracts points of interest according to a Heitger and Rosenthaler detector. In [14], they utilize a Harris detector to choose interest points in an image. An indexing system to better match images to a training set and a grouping of close-lying interest points to form ``semi-local'' features is introduced.

In Mikolajczyk and Schmid [15], a comparison is made of various methods of extracting points of interest and the efficacy of different region descriptors. In their experiments the matching vector utilized by Lowe are found to work best. Comparisons were made with various other descriptors including the affine invariant descriptor developed by them, steerable filters, cross correlations and differential invariants. Only in the face of illumination changes did the SIFT operator performance not achieve the best rating. This is an indication that methods that utilize histograms to describe features work well.

IV. EXPERIMENTS IN SCENE RECOGNITION

In this article we attempt to use the image matching strategy developed by Lowe. It consists of 1) feature extraction, 2) feature description through the construction of a large vector description 3) Matching using a modified K-d tree for a speedy search and matching of features. We have adopted the theory for discrete space developed by Lindeberg in the feature extraction stage. The creation of the descriptors of the features detected in the earlier stage is carried out using Lowe's method. The flowchart for the procedure use is shown in Figure 1. This 128 element vector describing the histogram of gradient orientations at 16 neighboring points of the point of interest is utilized. Each of these 128 values is given equal weights and a K-d tree with up-to 128 different dimensions is built to accommodate the database of interest features from all the models. As expressed by Lowe, this implementation should allow the collection of large numbers of points in a database for posterior matching while keeping the probability of erroneous detection low.

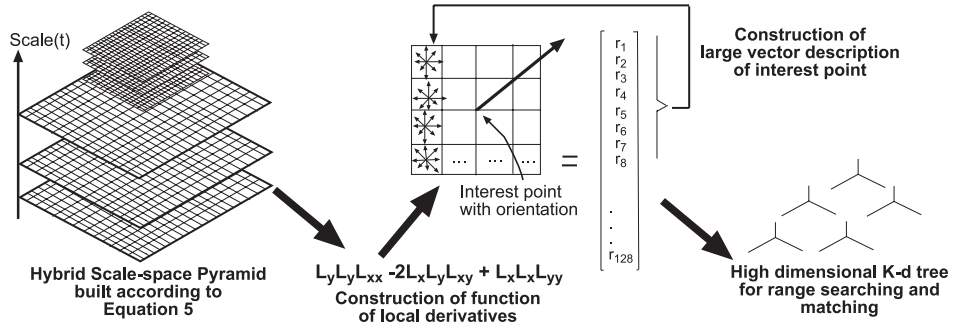

Figure 1. Pyramid Construction and point extraction follow from linear scale-space; Feature point description and matching are due to Lowe

The 128-long description vector was created in a similar way to that described by Lowe in [16]. At this stage we have attempted to match only the 100 best points (strongest Junction response) of pairs of images.

Two sets of images one taken in the exterior in bright sun light and the other in the interior of a building, were used. These are represented in the first row of Figure 2 and Figure 3. The algorithm for extraction of points and the creation of a vector description was run on images represented in the second and third rows, the latter taken from a different viewpoint from the former. The images are therefore subject to scaling and a general transformation. An ordinary digital camera was utilized without calibration.

The results, shown in Figure 2, show that some 'matched' interest points in the right image has not been located at positions corresponding to their position in the right image (feature 14). The matching results are affected by the presence of multiple similar geometry features (feature marked 1, 8, 16 and 24 for example in Figure 3) in the images resulting in some ill-matching, since no location constraints has been applied during the matching procedure. In both sets of images we notice the scale-space behavior of the points in which the corresponding in the zoomed image are detected at higher scales, equation (4).

V. EXTENSION TO 3-D

Our particular aim in studying multi-scale image analysis is to achieve a description, of 3-D objects. Whilst the use of illumination is widespread and has its advantages, the presence of occluding edges and an uncontrolled blurring of different objects results in unpredictable and complex behavior of the features in scale-space. Utilizing depth and object discontinuities could provide an enhanced description of objects for navigational applications.

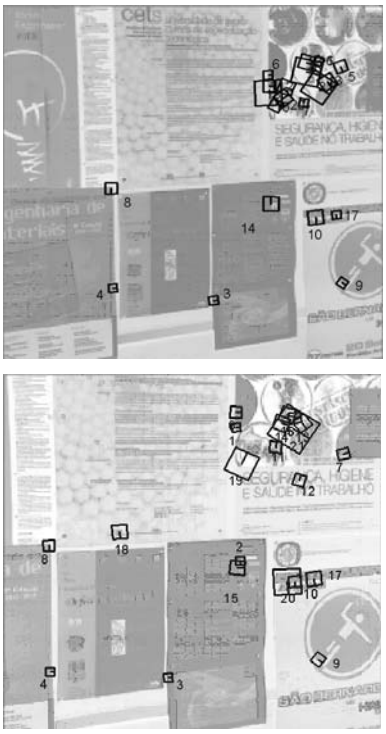

Figure 2. Initial scene taken in artificial lighting conditions

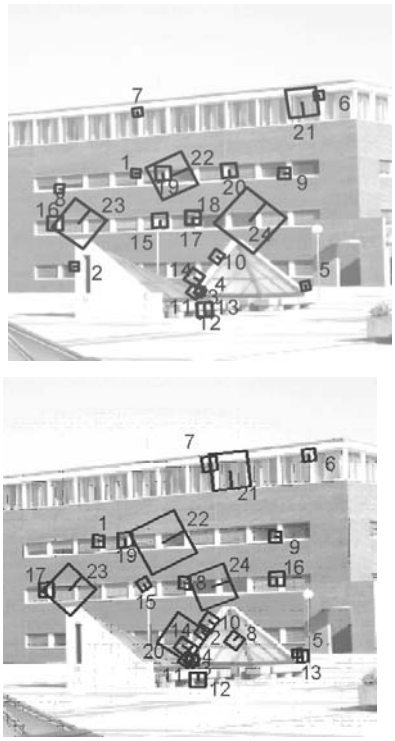

Figure 3. Exterior scene with a building having a number of similar architectural features

There has been some development of such techniques for methods that provide more information besides the two dimensional image. In extensions of the application of the SIFT method [17], matching across stereo views is used to increases the robustness of the selected features to changes in the viewpoint. Methods have been developed that extend the use of scale-space analysis in 2-D illumination images to problems in which sensors provide 3-D data. In attempts to describe the position, size and relationship between regions of blood circulation in the brain such methods have been developed for 3-D tomography data [18]. Attempts have also been made to use the method on 3-D seismic data.

We have extended the method of multi-scale image analysis to select SIFT features in regions with gradual change in depth, selected by directly using the values of the image disparity obtained from the MEGA-D stereo head by Videre Design and the stereo matching and 3-D reconstruction application, Small Vision System (SVS), by SRI International [19].

Although, for a reasonable confidence level, the depth information expressed in the image coordinates of the reference camera is incomplete, this problem can be minimized since the stereo matching algorithm based on a correlation window works best on regions with texture and contrasting intensity. These are exactly the points that we shall select in a prior scale-space feature extraction phase, significantly reducing the problem that the incomplete disparity data presents.

Some of the advantages we expect to obtain from this approach are:

1. The SIFT-like descriptors are limited to regions of gradual variation in the normal to the surface and to regions of predominantly small variations in depth. Features are thus predominantly obtained on the surface of individual objects and not from the interaction of different objects and their respective backgrounds.
2. It can be argued that feature extraction might be robust to viewpoint changes if the features arise from regions at similar depths. It should also make a training phase in an object recognition application more robust, and practicable for a greater range of objects that cannot be physically isolated from their environment.

The procedure consists in obtaining the values of a and c that satisfy the equation (5) for each of the points (coordinates u and v) within the search window. The consistency of the estimated values of a and c for all of the points within the window are then verified using a Hough Transform.

$$disparity(u, v) = a(u) + c \tag{5}$$

The flowchart describing the algorithm that uses the disparity information is shown in Figure 4.

VI. FILTERING OF FEATURES USING DISPARITY

Experiments were performed on a set of stereo images of scenes that include the same objects taken from different viewpoints and orientations. The objects were placed relatively close to the camera in order to obtain a sufficiently complete and accurate disparity values.

The pictures are taken with a MEGA-D stereo camera after calibration and stereo matching was done using the Smallv stereo reconstruction application. The images represent a different viewpoints and different orientation of the objects in the image.

Results are shown for a few of the features that were matched. The creation of feature descriptions and their matching was carried out in the way described in the section ‘Experiments in scene recognition’.

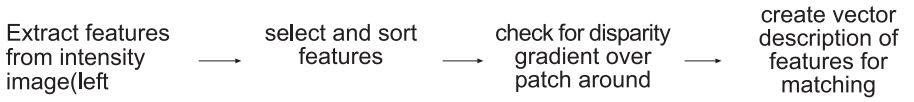

Figure 4. Using disparity to filter scale space features

Figure 5. Matched features after filtering using disparity information

VII. CONCLUSIONS AND FUTURE WORK

The extraction of features according to the hybrid scale-space method and their characterization using long descriptors presents good results. Other matching techniques are being attempted including matching of histogram distributions and the creation of descriptors that adjust to the scale of the image (more robust to changes in the zoom).

The inclusion of depth data from stereo allows for a better extraction of feature points. We expect the addition of this last phase to allow for improvements in the invariance of the features detected with changes in the viewpoint and a segmentation of detected features according to the objects in the scene that generated them.

Also the stereo information could be utilized to create a 3-D description of the object using many features for use in navigation and object recognition.

5. REFERENCES

[1] Cordelia Schmid and Roger Mohr, Matching by local invariants, Technical report, INRIA, 1995.

[2] Jan J Koenderink, The structure of images, Biological Cybernetics, 50(5), 1984.

[3] A.P.Witkin, Scale-space filtering, In Proc. 8th Joint conference on Artificial Intelligence, pages 1019--1023, Karlsruhe, W. Germany, 1983. 8th Joint conference on Artificial Intelligence.

[4] Tony Lindeberg, Scale-Space Theory in Computer Vision. Kluwer Academic Press, 1994.

[5] Bart M ter. Haar Romeny, editor, Geometry-Driven Diffusion in Computer Vision, Kluwer Academic Press, 1994.

[6] Tony Lindeberg, Scale-space for discrete signals, IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-12(3):234--254, 1990.

[7] Tony Lindeberg, Discrete derivative approximations with scale-space properties: A basis for low-level feature extraction. Journal of Mathematical Imaging and Vision, JMIV}, 3(3):349--376, 1993.

[8] Tony Lindeberg, Edge detection and ridge detection with automatic scale selection, Technical report, 1996.

[9] David G. Lowe, Object recognition from local scale-invariant features, In Proc. of the International Conference on Computer Vision, Corfu}, pages 1150--1157, 1999.

[10] M.Brown and Lowe D, Invariant features from interest point groups, In British Machine Vision Conference, BMVC 2002}, Cardiff, 2002.

[11] David G. Lowe, Local feature view clustering for 3d object recognition, Kauai, Hawaii, 2001.

[12] Arthur Pope and David G. Lowe, Probabilistic models of appearance for 3-d object recognition, IJCV}, (40,2), 2000.

[13] Brown M. and Lowe G. David, Recognizing panoramas, In Tenth International Conference on Computer Vision (ICCV 2003)}, October 2003.

[14] Cordelia Schmid and Roger Mohr, Local grayvalue invariants for image retrieval, IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(5):530--535, 1997.

[15]Krystian Mikolajczyk and Cordelia Schmid, Indexing based on scale invariant interest points, In International Conference on Computer Vision, Vancouver, 2001.

[16] David G. Lowe, Distinctive image features-from scale-invariant keypoints, International Journal of Computer Vision, 2003.

[17] Stephen Se, David Lowe, and Jim Little, Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks, International Journal of Robotics Research, Volume 21, Number 8, pages 735-758, August 2002.

[18] O.Coulon, I .Bloch, V.Frouin, and J-F. Mangin, Multiscale measures in linear scale-space for characterizing cerebral functional activations in 3d pet difference images, In ScaleSpace'97, First International Conference on Scale-Space Theory in Computer Vision, page 188.

[19] www.ai.sri.com/software/svs.