# Local Greyvalue Invariants for Image Retrieval

Cordelia Schmid and Roger Mohr

*Abstract*— **This paper addresses the problem of retrieving images from large image databases. The method is based on local greyvalue invariants which are computed at automatically detected interest points. A voting algorithm and semi-local constraints make retrieval possible. Indexing allows for efficient retrieval from a database of more than 1000 images. Experimental results show correct retrieval in the case of partial visibility, similarity transformations, extraneous features, and small perspective deformations.**

*Keywords*— **image retrieval, image indexing, greylevel invariants, matching**

## I. Introduction

This paper addresses the problem of matching an image to a large set of images. The query image is a new (partial) image of an object imaged in the database. The image may be taken from a different viewing angle (made precise below).

### A. Existing recognition methods

Existing approaches in the literature are of two types: those that use geometric features of an object; and those that rely on the luminance signature of an object.

Geometric approaches model objects by 3D properties such as lines, vertices and ellipses and try to extract these features in order to recognise the objects. General surveys on such model-based object recognition systems are presented in [1], [2]. These methods generally comprise three components: matching, pose computation, and verification. The key contribution of several recognition systems has been a method of cutting down the complexity of matching. For example tree search is used in [3] and recursive evaluation of hypotheses in [4]. In indexing, the feature correspondence and search of the model database are replaced by a look-up table mechanism [5], [6], [7]. The major difficulty of these geometry based approaches is that they use human-made models or require CAD-like representations. These representations are not available for objects such as trees or paintings; in the case of "geometric" objects these CAD-like representations are difficult to extract from the image.

An alternative approach is to use the luminance information of an object. The idea is not to impose what has to be seen in the image (points, lines ...) but rather to use what is really seen in the image to characterise an object. The first idea was to use colour histograms [8]. Several authors have improved the performance of the original colour histogram matching technique by introducing measures which are less sensitive to illumination changes [9], [10], [11], [12]. Instead of using colour, greyvalue descriptors can also be used for histograms [13]. Another idea is to use a collection of images and reduce them in an eigenspace. This approach was first used in [14] for face recognition and then in [15] for general objects. A different reduction is proposed in [16] who learns features which best describe the image. It is also possible to compute local greyvalue descriptors at points of a global grid. The descriptors are either steerable filters [17] or Gabor filters [18], [19]. In the case of partial visibility grid placement gets difficult, as the grid cannot be centred.

### B. Our approach

All of the existing luminance approaches are global and therefore have difficulty in dealing with partial visibility and extraneous features. On the other hand, geometric methods have difficulties in describing "non-geometric" objects and they have problems differentiating between many objects. Local computation of image information is necessary when dealing with partial visibility; photometric information is necessary when dealing with a large number of similar objects. The approach described here uses local greyvalue features computed at interest points as displayed in figure 1. Interest points are local features with high informational content (section II).

The local characteristics used in this work are based on differential greyvalue invariants [20], [21]. This ensures invariance under the group of displacements within an image. A multi-scale approach [22], [23] makes this characterisation robust to scale changes, that is to similarity transformations (section III). Due to a stable implementation of these invariants, a reliable characterisation of the signal is obtained. Moreover, this characterisation is significant, as it is computed at interest points.

A voting algorithm makes retrieval robust to miss-matches as well as outliers. Outliers are caused by miss-detection of feature points and extraneous features. Semi-local constraints reduce the number of miss-matches. Furthermore, indexing via a multi-dimensional hash-table makes fast retrieval possible (section IV).

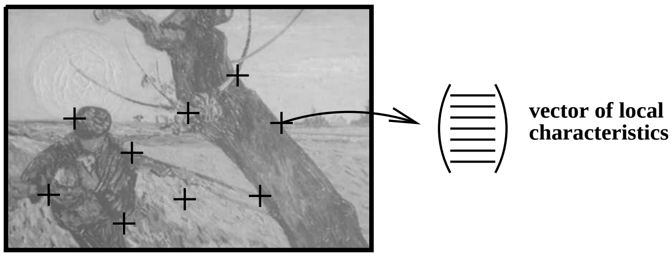


Fig. 1. Representation of an image.

Our approach allows the handling of partial visibility and transformations such as image rotation and scaling (section V). Experiments have been conducted on a set of more than a thousand images, some of them very similar in shape or texture. The high recognition rate is the result of careful design in which robustness to outliers and tolerance to image noise were considered at each step.

The authors are with GRAVIR-IMAG and INRIA Rhône-Alpes, 655 avenue de l'Europe, 38330 Monbonnot Saint-Martin, France. E-mail : cordelia.schmid@inrialpes.fr.

Cordelia Schmid has been partially supported by the HCM program of the European Community.

## II. Interest Points

Computing image descriptors for each pixel in the image creates too much information. Interest points are local features at which the signal changes two-dimensionally. The use of interest points has advantages over features such as edges or regions, particularly robustness to partial visibility and high informational content.

A wide variety of detectors for interest points exists in the literature, the reader is referred to [24] for an exhaustive overview. In the context of matching, detectors should be repeatable. A comparison of different detectors under varying conditions [25] has shown that most repeatable results are obtained for the detector of Harris [26]. The basic idea of this detector is to use the auto-correlation function in order to determine locations where the signal changes in two directions. A matrix related to the auto-correlation function which takes into account first derivatives of the signal on a window is computed :

$$\exp^{-\frac{x^2+y^2}{2\sigma^2}} \otimes \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$$

The eigenvectors of this matrix are the principal curvatures of the auto-correlation function. Two significant values indicate the presence of an interest point.

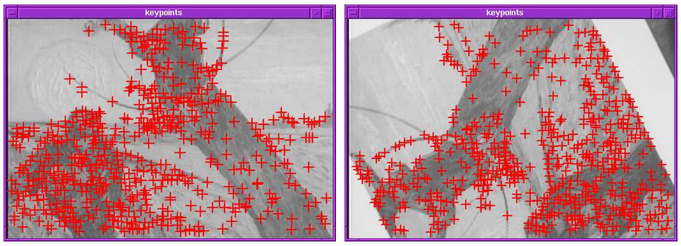


Fig. 2. Interest Points detected on the same scene under rotation. The image rotation between the left image and the right image is 155 degrees. The repeatability rate is 92%.

Figure 2 shows interest points detected on the same scene under rotation. The repeatability rate is 92% which means that 92% of the points detected in the first image are detected in the second one. Experiments with images taken under different conditions show that the average repeatability rate is about 90%. Moreover, 50% repeatability is sufficient for the remaining process if we use robust methods.

## III. Multi-scaled differential greyvalue invariants

Our characterisation is based on derivatives which locally describe an image. In order to obtain invariance under the group $SO(2)$ of rigid displacements in the image, differential invariants are computed. These invariants are then inserted into a multi-scale framework in order to deal with scale changes. Therefore the characterisation is invariant to similarity transformations which are additionally quasi-invariant to 3D projection (see [27]).

### A. Local jet

The image in a neighbourhood of a point can be described by the set of its derivatives. Their stable computation is achieved by convolution with Gaussian derivatives [28], [22], [23]. This set of derivatives has been named "local jet" by Kœnderink [20] and defined as follows:

Let $I$ be an image and $\sigma$ a given scale. The "local jet" of order $N$ at a point $\mathbf{x} = (\mathbf{x_1}, \mathbf{x_2})$ is defined by

$$J^N[I](\mathbf{x},\sigma) = \{L_{i_1 \ldots i_n}(\mathbf{x},\sigma) \mid (\mathbf{x},\sigma) \in I \times I\!\!R^+ ; n = 0, \ldots, N\}$$

in which $L_{i_1 \ldots i_n}(\mathbf{x},\sigma)$ is the convolution of image $I$ with the Gaussian derivatives $G_{i_1 \ldots i_n}(\mathbf{x},\sigma)$ and $i_k \in \{x_1, x_2\}$.

The $\sigma$ of the Gaussian function determines the quantity of smoothing. This $\sigma$ also coincides with a definition of scale-space which will be important for our multi-scale approach. In the following, $\sigma$ will be referred to as the *size* of the Gaussian.

### B. Complete set of differential invariants

In order to obtain invariance under the group $SO(2)$, differential invariants are computed from the local jet. Differential invariants have been studied theoretically by Kœnderink [20] and Romeny et al.[28], [29], [30]. A complete set of invariants can be computed that locally characterises the signal. The set of invariants used in this work is limited to third order. This set is stacked in a vector, denoted by $\mathcal{V}$. In equation 1 vector $\mathcal{V}$ is given in tensorial notation – the so-called Einstein summation convention. Notice that the first component of $\mathcal{V}$ represents the average luminance, the second component the square of the gradient magnitude and the fourth the Laplacian.

$$\mathcal{V}[0..8] = \begin{bmatrix} L \\ L_i L_i \\ L_i L_{ij} L_j \\ L_{ii} \\ L_{ij} L_{ji} \\ \varepsilon_{ij}(L_{jkl} L_i L_k L_l - L_{jkk} L_i L_l L_l) \\ L_{iij} L_j L_k L_k - L_{ijk} L_i L_j L_k \\ -\varepsilon_{ij} L_{jkl} L_i L_k L_l \\ L_{ijk} L_i L_j L_k \end{bmatrix} \quad (1)$$

with $L_i$ being the elements of the "local jet" and $\varepsilon_{ij}$ the 2D antisymmetric Epsilon tensor defined by $\varepsilon_{12} = -\varepsilon_{21} = 1$ and $\varepsilon_{11} = \varepsilon_{22} = 0$.

### C. Multi-scale approach

To be insensitive to scale changes the vector of invariants has to be calculated at several scales. A methodology to obtain such a multi-scale representation of a signal has been proposed in [31], [22], [23].

For a function $f$, a scale change $\alpha$ can be described by a simple change of variables, $f(x) = g(u)$ where $g(u) = g(u(x)) = g(\alpha x)$. For the nth derivatives of f, we obtain $f^{(n)}(x) = \alpha^n g^{(n)}(u)$. Theoretical invariants are then easy to derive, for example $\frac{\left[f^{(n)}(x)\right]^{\frac{k}{n}}}{f^{(k)}(x)}$ is such an invariant.

However, in the case of a discrete representation of the function, as for an image, derivatives are related by :

$$\int_{-\infty}^{+\infty} I_1(\vec{x}) G_{i_1 \ldots i_n}(\vec{x}, \sigma) d\vec{x} = \alpha^n \int_{-\infty}^{+\infty} I_2(\vec{u}) G_{i_1 \ldots i_2}(\vec{u}, \sigma\alpha) d\vec{u} \quad (2)$$

with $G_{i_1 \ldots i_2}$ being the derivatives of the Gaussian.

Equation 2 shows that the size of the Gaussian has to be adjusted which implies a change of the calculation support. As it is impossible to compute invariants at all scales, scale quantisation is necessary for a multi-scale approach. Often a half-octave quantisation is used. The stability of the characterisation has proven this not to be sufficient. Experiments have shown that matching based on invariants is tolerant to a scale change of 20% (see [25]). We have thus chosen a scale quantisation which ensures that the difference between consecutive sizes is less than 20%. As we want it to be insensitive to scale changes up to a factor of 2, the size $\sigma$ varies between 0.48 and 2.07, its values being: 0.48, 0.58, 0.69, 0.83, 1.00, 1.20, 1.44, 1.73, 2.07.

## IV. Retrieval Algorithm

To retrieve an image, it is necessary to decide if two invariant vectors are similar. Similarity is quantified using the Mahalanobis distance. To define the distance for a set of vectors which includes outliers to the database a voting algorithm has to be used. An indexing technique make access fast ; and semi-local constraints allow to reduce miss-matches.

### A. Vector comparison by Mahalanobis distance

A standard method is to model the uncertainties in the components as random variables with Gaussian distribution and use the Mahalanobis distance $d_M$ to compare invariant vectors. This distance takes into account the different magnitude as well as the covariance matrix $\Lambda$ of the components. For two vectors $\mathbf{a}$ and $\mathbf{b}$, $d_M(\mathbf{b}, \mathbf{a}) = \sqrt{(\mathbf{b}-\mathbf{a})^T \Lambda^{-1} (\mathbf{b}-\mathbf{a})}$.

The square of the Mahalanobis distance is a random variable with a $\chi^2$ distribution. Since the square root function is a bijection from $I\!R^+$ to $I\!R^+$, it is possible to use a table of this distribution to threshold the distance and then to reject the k% of values that are most likely to correspond to false matches.

In order to obtain accurate results for the distance, it is important to have a representative covariance matrix which takes into account signal noise, luminance variations as well as imprecision of the interest point location. As a theoretical computation seems impossible to derive given realistic hypotheses, we estimated it statistically by tracking interest points in image sequences.

The Mahalanobis distance is impractical for implementing a fast indexing technique. However, a base change makes conversion into the standard Euclidean distance $d_E$ possible. As the covariance matrix is a real symmetric (semi) definite positive matrix, it can be decomposed into $\Lambda^{-1} = P^T D P$ where P is orthogonal and D is diagonal. We then have $d_M(\mathbf{a}, \mathbf{b}) = d_E(\sqrt{D} P \mathbf{a}, \sqrt{D} P \mathbf{b})$.

### B. Indexing and voting algorithm

*B.1. Voting algorithm.* A database contains a set $\{M_k\}$ of models. Each model $M_k$ is defined by the vectors of invariants $\{\mathcal{V}_j\}$ calculated at the interest points of the model images. During the storage process, each vector $\mathcal{V}_j$ is added to the database with a link to the model $k$ for which it has been computed. Formally, the simplest database is a table of couples $(\mathcal{V}_j, k)$.

Recognition consists of finding the model $M_{\hat{k}}$ which corresponds to a given query image $I$, that is the model which is most similar to this image. For this image a set of vectors $\{\mathcal{V}_l\}$ is computed which corresponds to the extracted interest points. These vectors are then compared to the $\mathcal{V}_j$ of the base by computing: $d_M(\mathcal{V}_l, \mathcal{V}_j) = d_{l,j} \ \forall (l, j)$. If this distance is below a threshold $t$ according the $\chi^2$ distribution, the corresponding model gets a vote.

As in the case of the Hough transform [32], the idea of the voting algorithm is to sum the number of times each model is selected. This sum is stored in the vector $T(k)$. The model that is selected most often is considered to be the best match: the image represents the model $M_{\hat{k}}$ for which $\hat{k} = \arg\ \max_k T(k)$.

Figure 3 shows an example of a vector $T(k)$ in the form of a histogram. Image 0 is correctly recognised. However, other images have obtained almost equivalent scores.

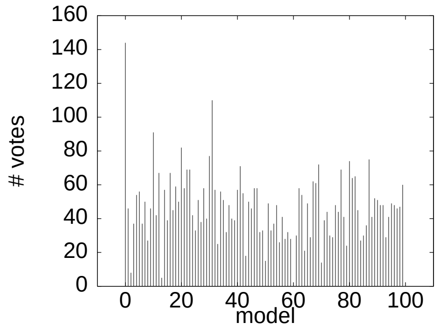


Fig. 3. Result of the voting algorithm: the number of votes are displayed for each model image. Image 0 is recognised correctly.

*B.2. Multi-dimensional indexing.* Without indexing the complexity of the voting algorithm is of the order of $l \times N$ where $l$ is the number of features in the query image and $N$ the total number of features in the data base. As $N$ is large (about 150,000 in our tests) efficient data structures need to be used.

Search structures have been extensively studied. An overview of all tree-like data structures that allow fast and/or compact access to data is presented in [33]. The data structure used here is not referenced in the previous review; it can be seen as a variant of $k$-d trees.

Here each dimension of the space is considered sequentially. Access to a value in one dimension is made through fixed size 1-dimensional buckets. Corresponding buckets and their neighbours can be directly accessed. Accessing neighbours is necessary to take into account uncertainty. A bucket is extended in the next dimension if the number of values stored is above a threshold. Therefore the data structure can be seen as a tree with a depth which is at most the number of dimensions of the stored vectors. The complexity of indexing is of the order of l (number of features of the query image).

This indexing technique leads to a very efficient recognition. The database contains 154030 points. The mean retrieval time for our database containing 1020 objects is less than 5 seconds on a Sparc 10 Station. Performance could be further improved by parallelisation, as each vector is processed separately.

### C. Semi-local constraints

A given feature might vote for several models. Having a large number of models or many very similar ones raises the probability that a feature will vote for several models. Califano [34] suggested that using longer vectors decreases this probability. Yet the use of higher order derivatives for our invariants is not practical. Another way to decrease the probability of false matches is to use global features. However, global characteristics are sensitive to extraneous features and partial visibility.

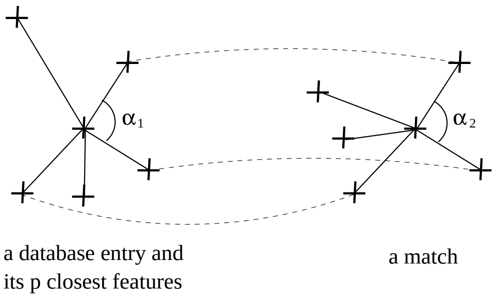


Fig. 4. Semi-local constraints : neighbours of the point have to match and angles have to correspond. Note that not all neighbours have to be matched correctly.

Our solution is the use of local shape configurations, as in figure 4. Semi-local constraints have previously been used in [35], [36]. For each feature (interest point) in the database, the $p$ closest features in the image are selected. If we require that all $p$ closest neighbours are matched correctly, we suppose that there is no miss-detection of points. Therefore, we require that at least 50% of the neighbours match. In order to increase the recognition rate further, a geometric constraint is added. This constraint is based on the angle between neighbour points. As we suppose that the transformation can be locally approximated by a similarity transformation, these angles have to be locally consistent, for example the angles $\alpha_1$ and $\alpha_2$ in figure 4. An example using the geometrical coherence and the semi-local constraints is displayed in figure 5. It gives the votes if constraints are applied to the example in figure 3. The score of the object to be recognised is now much more distinctive.

### D. Multi-scale approach

The multi-scale approach can be very easily integrated into the framework presented above. For a query image, invariants are computed at several scales (see section III).

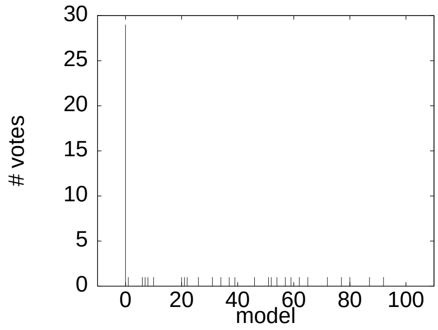


Fig. 5. Result of applying semi-local constraints : the number of votes are displayed for each model image. Semi-local constraints decrease the probability of false votes. Image 0 is recognised much more distinctively than in figure 3.

Invariants in the database are only stored for one scale. Matching invariants computed at several scales to invariants computed at one scale increases the possibility of wrong matches and makes semi-local constraints even more essential. These constraints implicitly include a scale constraint, as the invariants for a point and its neighbours are calculated at the same scale. Hence, if a point and its neighbours match, the scale constraint is fulfilled. Thus, using these constraints, the multi-scale approach works efficiently as is demonstrated in the next section.

## V. Experimental Results

Experiments have been conducted for an image database containing 1020 images. They have shown the robustness of the method to image rotation, scale change, small viewpoint variations, partial visibility and extraneous features. The obtained recognition rate is above 99% for a variety of test images taken under different conditions.

### A. Content of the database

The database includes different kinds of images such as 200 paintings, 100 aerial images and 720 images of 3D objects (see figure 6). 3D objects include the Columbia database. These images are of a wide variety. However, some of the painting images and some of the aerial images are very similar. This leads to ambiguities which the recognition method is capable of dealing with.

In the case of a planar 2D object, an object is represented by one image in the database. This is also the case for nearly planar objects as for aerial images. A 3D object has to be represented by images taken from different viewpoints. Images are stored in the database with 20 degrees viewpoint changes.

### B. Recognition results

In this section some examples illustrate the different conditions under which the method can still operate correctly. A systematic evaluation for a large number of test images taken under different conditions is then presented. More details are given in [25].

*B.1. Some examples of correct recognition.* In the following three examples are displayed, one for each type of

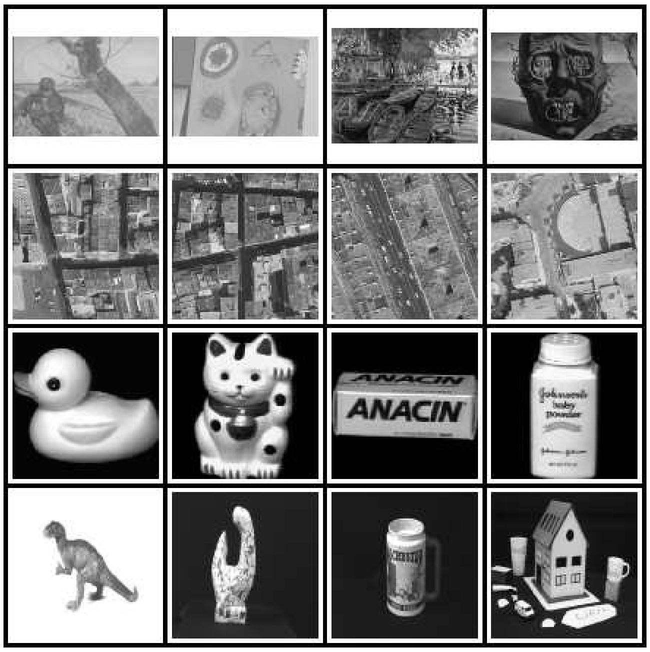


Fig. 6. Some images of the database. The database contains more 1020 images.

image. For all of them, the image on the right is stored in the database. It is correctly retrieved using any of the images on the left. Figure 7 shows recognition of a painting image in the case of image rotation and scale change. It also shows that correct recognition is possible if only part of an image is given.

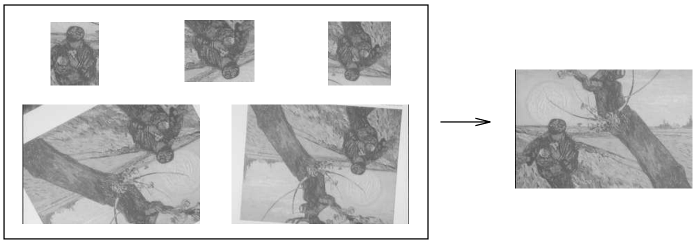

Fig. 7. The image on the right is correctly retrieved using any of the images on the left. Images are rotated, scaled and only part of the image is given.

In figure 8 an example of an aerial image is displayed. It shows correct retrieval in the case of image rotation and if part of an image is used. In the case of aerial images we also have to deal with a change in viewpoint and extraneous features. Notice that buildings appear differently because viewing angles have changed and cars have moved.

Figure 9 shows recognition of a 3D object. The object has been correctly recognised in the presence of rotation, scale change, change in background and partial visibility. In addition, there is a change of 10 degrees of viewpoint position between the two observations. Notice that the image of the object has not only been recognised correctly, but that the closest stored view has also been retrieved.

*B.2. Systematic evaluation of retrieval.* The method is evaluated for different transformations – image rotation, scale change, viewpoint variations – as well as for partial visibility.

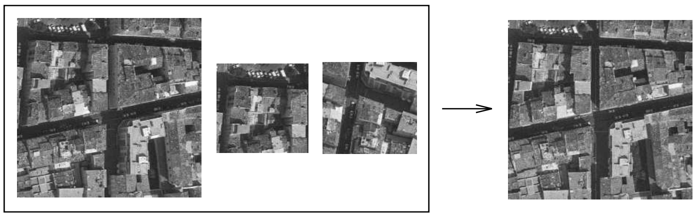

Fig. 8. The image on the right is correctly retrieved using any of the images on the left. Images are seen from a different viewpoint (courtesy of Istar).

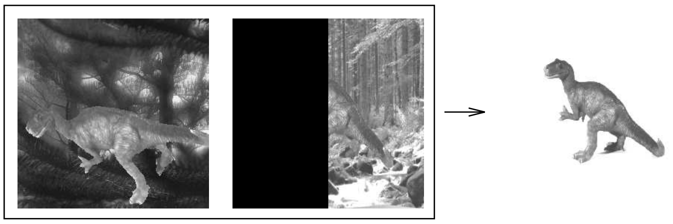

Fig. 9. The image on the right is correctly retrieved using any of the images on the left. The 3D object is in front of a complex background and only partially visible.

*Image rotation* To test invariance to image rotation, images were taken by rotating the camera around its optical axis. The recognition rate obtained is 100% for different rotations equally distributed over a circle. This experiment shows that the characterisation is completely invariant to image rotation.

*Scale change* Experiments were conducted on zoomed images. Using a multi-scale approach, the recognition rate attains a score of 100% up to a scale change of 2. At present, this factor seems to be the limit for our method. However, this limit is not due to our invariant characterisation but rather to the stability of the interest point detector. The repeatability of this detector decreases rapidly when the scale change is greater than 1.6.

*Viewpoint variation* Test images are taken at angles different from the images stored in the base. Each aerial image has been taken from 4 different viewpoints. Viewpoint number one is stored in the base. For images taken from different viewpoints, the recognition rate is 99%. The only image which is not recognised correctly is part of the harbour and contains only water on which no reliable interest points can be identified.

For 3D objects, test images have been taken at 20 degrees difference in viewing angle. The viewing angles of the test images lie in between two images stored in the base. The recognition rate is 99.86%. It is interesting to consider only the Columbia database which serves as a benchmark for object recognition. On this base a 100% recognition rate has been obtained in [15] as well as in [17]. Experiments show that our method attains the same recognition rate.

*Partial visibility* Parts of different size are extracted randomly from painting images. The relative size varies between 10% and 100%. For parts of relative size greater than or equal to 30%, the recognition rate is 100%. For a relative size of 20%, a 95% rate is obtained; and for a relative size of 10%, a 90% rate. Considering the size of our database, this can be explained by the fact that points are very discriminating and thus only a few points are necessary to recognise an image. It is thus possible to retrieve an image even if only part of this image is given. However, very small parts do not contain enough points, so the number of votes is limited. In this case, the robust algorithm can not overcome the statistical uncertainty.

## VI. Conclusion

This paper has shown that the differential greylevel invariants introduced by Kœnderink efficiently characterise points. These invariants describe the image locally. As automatically detected interest points are characteristics of patterns, invariants calculated at interest points can be used for indexing 2D greylevel patterns. A voting algorithm and multi-dimensional indexing make image retrieval possible. However, blindly voting on individual invariants is not sufficient to guarantee the correctness of the answer in database indexing. It is then crucial to introduce a semi-local coherence between these identifications. This increases the recognition rate. Experiments conducted on a database containing 1020 images have shown very good results. Even small parts of images can be recognised correctly. This is due to the fact that the proposed characterisation is very discriminating.

Finally, different extensions are possible. The voting algorithm can be improved by taking into account the statistical distributions of the invariants; some of the invariants are more discriminating than others. In addition, computation of a confidence value is then possible.

Using global consistency checking for local matches or a global constraint such as the epipolar geometry is another possible extension. Such additional constraints further increase the recognition rate and make detection of several objects possible.

## References


[1] P.J. Besl and R.C. Jain, "Three-dimensional object recognition," *ACM Computing Surveys*, vol. 17, no. 1, pp. 75–145, 1985.

[2] R.T. Chin, H. Smith, and S.C. Fralick, "Model-based recognition in robot vision," *ACM Computing Surveys*, vol. 18, no. 1, pp. 67–108, 1986.

[3] R.C. Bolles and R. Horaud, "3DPO : A three-dimensional Part Orientation system," *Int'l J. Robotics Research*, vol. 5, no. 3, pp. 3–26, 1986.

[4] N. Ayache and O.D. Faugeras, "HYPER: a new approach for the recognition and positioning of 2D objects," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, no. 1, pp. 44–54, 1986.

[5] D.J. Clemens and D.W. Jacobs, "Model-group indexing for recognition," in *Proc. DARPA Image Understanding Workshop*, 1990, pp. 604–613.

[6] Y. Lamdan and H.J. Wolfson, "Geometric hashing: a general and efficient model-based recognition scheme," in *Proc. 2nd Int'l Conf. on Computer Vision*, 1988, pp. 238–249.

[7] C.A. Rothwell, *Object Recognition Through Invariant Indexing*, Oxford Science Publications, 1995.

[8] M.J. Swain and D.H. Ballard, "Color indexing," *Int'l J. of Computer Vision*, vol. 7, no. 1, pp. 11–32, 1991.

[9] B.V. Funt and G.D. Finlayson, "Color constant color indexing," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 17, no. 5, pp. 522–529, 1995.

[10] K. Nagao, "Recognizing 3D objects using photometric invariant," in *Proc. 5th Int'l Conf. on Computer Vision*, 1995, pp. 480–487.

[11] S.K. Nayar and R.M. Bolle, "Computing reflectance ratios from an image," *Pattern Recognition*, vol. 26, no. 10, pp. 1529–1542, 1993.

[12] D. Slater and G. Healey, "The illumination-invariant recognition of 3D objects using color invariants," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 2, pp. 206–210, 1996.

[13] B. Schiele and J.L. Crowley, "Object recognition using multidimensional receptive field histograms," in *Proc. 4th European Conf. on Computer Vision*, 1996, pp. 610–619.

[14] M.A. Turk and A.P. Pentland, "Face recognition using eigenfaces," in *Proc. Conf. on Computer Vision and Pattern Recognition*, 1991, pp. 586–591.

[15] H. Murase and S.K. Nayar, "Visual learning and recognition of 3D objects from appearance," *Int'l J. of Computer Vision*, vol. 14, pp. 5–24, 1995.

[16] P. Viola, "Feature-based recognition of objects," in *Proceedings of the AAAI Fall Symposium Series: Machine Learning in Computer Vision: What, Why, and How?*, 1993, pp. 60–64.

[17] R.P.N. Rao and D.H. Ballard, "An active vision architecture based on iconic representations," *Artificial Intelligence*, pp. 461–505, 1995.

[18] M. Lades, J.C. Vorbrüggen, J. Buhmann, J. Lange, C.v.d. Malsburg, R.P. Würtz, and W. Konen, "Distortion invariant object recognition in the dynamic link architecture," *IEEE Trans. Computers*, vol. 42, no. 3, pp. 300–311, 1993.

[19] X. Wu and B. Bhanu, "Gabor wavelets for 3D object recognition," in *Proc. 5th Int'l Conf. on Computer Vision*, 1995, pp. 537–542.

[20] J. J. Koenderink and A. J. van Doorn, "Representation of local geometry in the visual system," *Biological Cybernetics*, vol. 55, pp. 367–375, 1987.

[21] B.M. ter Haar Romeny, *Geometry-Driven Diffusion in Computer Vision*, Kluwer Academic Publishers, 1994.

[22] T. Lindeberg, *Scale-Space Theory in Computer Vision*, Kluwer Academic Publishers, 1994.

[23] A.P. Witkin, "Scale-space filtering," in *Int'l Joint Conf. on Artifical Intelligence*, 1983, pp. 1019–1023.

[24] R. Deriche and G. Giraudon, "A computational approach for corner and vertex detection," *Int'l J. of Computer Vision*, vol. 10, no. 2, pp. 101–124, 1993.

[25] C. Schmid, *Appariement d'images par invariants locaux de niveaux de gris*, Ph.D. thesis, Institut National Polytechnique de Grenoble, 1996.

[26] C. Harris and M. Stephens, "A combined corner and edge detector," in *Alvey Vision Conf.*, 1988, pp. 147–151.

[27] T.O. Binford and T.S. Levitt, "Quasi-invariants: Theory and exploitation," in *Proc. DARPA Image Understanding Workshop*, 1993, pp. 819–829.

[28] Luc Florack, *The Syntactical Structur of Scalare Images*, Ph.D. thesis, Universiteit Utrecht, 1993.

[29] L.M.T. Florack, B. ter Haar Romeny, J.J Koenderink, and M.A. Viergever, "General intensity transformation and differential invariants," *J. Mathematical Imaging and Vision*, vol. 4, no. 2, pp. 171–187, 1994.

[30] B.M. ter Haar Romeny, L.M.J. Florack, A.H. Salden, and M.A. Viergever, "Higher order differential structure of images," *Image and Vision Computing*, vol. 12, no. 6, pp. 317–325, 1994.

[31] J.J. Koenderink, "The structure of images," *Biological Cybernetics*, vol. 50, pp. 363–396, 1984.

[32] S.D. Shapiro, "Feature space transforms for curve detection," *Pattern Recognition*, vol. 10, no. 3, pp. 129–143, 1978.

[33] A. Samet, "The quadtree and related hierarchical data structures," *ACM Computing Surveys*, vol. 16, no. 2, pp. 189–259, 1984.

[34] A. Califano and R. Mohan, "Multidimensional indexing for recognizing visual shapes," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 16, no. 4, pp. 373–392, 1994.

[35] L. van Gool, P. Kempenaers, and A. Oosterlinck, "Recognition and semi-differential invariants," in *Proc. Conf. on Computer Vision and Pattern Recognition*, 1991, pp. 454–460.

[36] Z. Zhang, R. Deriche, O. Faugeras, and Q.T. Luong, "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry," *Artificial Intelligence*, vol. 78, pp. 87–119, 1995.