

An Adaptive-Sampling Algorithm for Object Representation

Robert Alterson and Minas Spetsakis

Department of Computer Science, York University, Toronto, Ontario

{alterson,minas}@cs.yorku.ca

Abstract

We present a novel adaptive-sampling algorithm for spectral signature generation. This algorithm is designed to increase inter-object discrimination and reduce feature-vector dimensionality. Our algorithm is applied to a Gabor-feature based multi-resolutional object detection and recognition scheme. In this context we study and analyze the detection and identification of unknown objects in a complex background. Iterative, off-line optimization methods are employed to reduce computational demands during the learning phase. Our representation scheme takes into account all items in a given object library. It selects sample-point sets that maximize the inter-object distance. Thus, the presented method increases identification robustness and can reduce the size of signature vectors.

1 Introduction

This paper describes an algorithm for object representation that can be applied to several transform domains. Objects are sampled in order to extract signature vectors containing highly discriminant data. These vectors are then used to represent each item in a local object library. Such feature sets are required to be as unique to the sampled object as possible to allow for maximum differentiation from other items. Several face-recognition methods make use of *fiducial* points, located at face landmarks that have been found to provide effective characterization, e.g., the eyes and the ears [13]. However, for the more general problem of object recognition we do not know in advance what object points carry more significant information than others for the purpose of distinguishing between items. These points have to be determined without human intervention.

Due to the high time and space costs associated with processing high dimensionality data, it is necessary to represent transformed objects by as short as possible feature vectors. Typically, this reduction is performed by taking a few samples from each object. However, the research literature does not provide optimal or efficient methods for sample selection.

In this paper we present results obtained for the Gabor transform. However, other transforms/filters can be applied for extracting spectral-signature information, e.g., the Fourier transform and derivatives of the Gaussian [2].

Two classes of object representation methods are mentioned in the literature: Statistical representation of local-image data and fixed-grid sampling [12]. Statistical representation methods typically represent objects by low-order statistics of a portion of their transform. In [1] Gabor filtered images are represented by the average and the standard deviation computed for all pixels. Fixed-grid sampling is described in [8], where Gabor filters are used for handwritten numeral recognition. A grid of 64 points at regular intervals is used to sample Gabor-filtered objects. The grid size is dynamically adjusted to ensure it contains the entire object circumference. Dailey and Cottrell [7] use a 29x36 fixed grid to sample faces filtered with 40 different Gabor kernels for a total of 41760 features. Principle Component Analysis is then applied in order to reduce the dimensionality to 109 features. Gabor filters are used for facial expression recognition in [16], where feature vectors are created by sampling Gabor-filtered faces at predetermined fiducial points e.g., the eyes and chin.

Ben-Arie and Pandit use five scales and four orientations for object detection. They represent objects by a single sample point in Gabor space [3]. Chung et al [6] use 12 predefined fiducial points at five central frequencies and eight orientations for face recognition. PCA is later used for dimensionality reduction. Wiskott et al sample the Gabor transform magnitude at various face landmarks and store them in a graph [13].

Increasing the separation between object representations is of prime importance. Even an infinitesimal variation between feature vectors can make the difference between a correct and incorrect recognition where nearest-neighbor classification is employed. The sampling method presented in this paper is unique because it determines preferred sampling points by considering all objects in a given library. Sample points are automatically selected so that they increase inter-object differentiation and thus can reduce required feature-vector lengths.

2 The Gabor Transform

Gabor filters can efficiently localize frequency and orientation properties of an analyzed image. The general form of the complex Gabor filter is presented below [9]. It consists of a two-dimensional Gaussian function with standard deviation σ , that modulates a spatial sinusoid,

$$Gabor(x, y) = e^{-\frac{x^2+y^2}{2\sigma^2}} \cdot e^{j \cdot (\omega_x \cdot x + \omega_y \cdot y)} \quad (1)$$

where ω_x and ω_y represent the radial frequencies in the horizontal and vertical directions, respectively. By convolving a Gabor filter with image I we obtain the image W :

$$W(x, y) = \iint I(u, v) Gabor(x-u, y-v) dudv \quad (2)$$

Gabor filters have been applied to several areas of Computer Vision. These include texture segmentation, document analysis, edge detection, retina identification, target detection, fractal dimension measurement, image coding, and image representation [14].

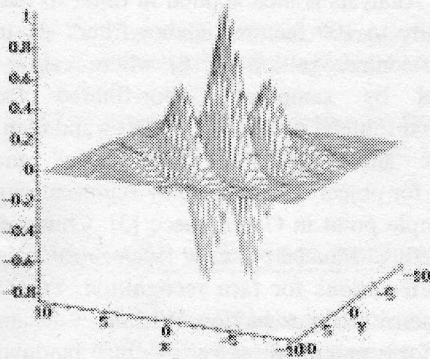


Figure 1: Spatial-domain even-symmetric 2D Gabor filter. This figure illustrates the real (even-symmetric) part of the filter. The central frequency here is 0.2 cycles per pixel (cpp).

3 A Dynamic Sampling Geometry

In order to adaptively sample transformed objects, a dynamic sampling structure is needed. The dynamic sampling geometry developed in this work features a higher concentration of sample points towards the centre and provides angular-coverage spacing of 45 degrees. A similar geometry has also been used in [4]. This geometry is represented mathematically by a central point and a set of surrounding point groups, with each group having eight members as follows:

$$G_d = \begin{cases} \{x_0, y_0\} & d = 0 \\ \{x_0 - d, y_0\}, \{x_0 - d, y_0 - d\}, \{x_0, y_0 - d\}, \\ \{x_0 + d, y_0 - d\}, \{x_0 + d, y_0\}, \{x_0 + d, y_0 + d\}, & d > 0 \\ \{x_0, y_0 + d\}, \{x_0 - d, y_0 + d\} \end{cases} \quad (3)$$

where G_d is the set of points corresponding to the spacing parameter d and $\{x_0, y_0\}$ is the location of the central point. The entire sampling geometry set is provided by:

$$G = \bigcup_{d=1}^F G_d \quad (4)$$

where the parameter F determines the number of point groups included in the geometry. The figure below illustrates three different sampling geometries for the F values, from left to right, of 2,3 and 4, respectively.

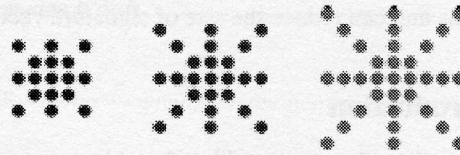


Figure 2: Sampling geometries for $F=2,3,4$.

The cardinality of the sampling geometry is therefore:

$$S = 1 + \sum_{i=1}^F 8 \cdot i \quad (5)$$

The geometry coverage span can be varied by linearly modulating the distance parameter d of each set G_d with a modulation constant, D having values $[1, 2, \dots, D_{max}]$. D_{max} is determined such that $2FD_{max}$, the maximum allowed span of the geometry, does not exceed the dimensions of the object image. Thus, each D value can produce a different sampling configuration. Note that non-adaptive sampling is merely a special case of the adaptive geometry with the modulation constant equal to 1. Figure 3 illustrates two sampling geometries, both with $F=2$ but with different modulation constants. The geometry on the left has a D value of 2 while the one on the right has a corresponding D value of 3.

Assuming our sampling geometry is always positioned on top of the centre of the object image, the problem is to find, for each object, a modulation constant that would maximize the inter-object distance value for the distance

metric of choice. Expanding or contracting the sample geometry is therefore analogous to increasing or decreasing the value of that constant, respectively. However, any object recognition method has to have some tolerance to changes in scale, orientation and et cetera. In our work, the continuous nature of the Gabor function helps a great deal in that respect.

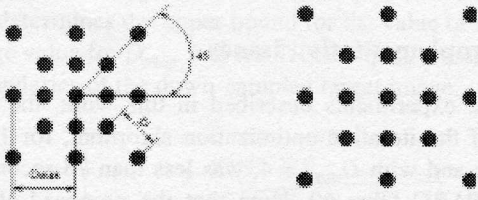


Figure 3: Sampling geometries ($F=2$) for two different modulation constants.

The outline of our approach to solving this optimization problem is as follows. For each object image in the database we compute its distance to the other objects. We repeat the latter computation for a series of possible optimization parameter values. The best sampling configuration for each object is selected to maximize the cumulative separation across all items. The object is subsequently represented in the database as a feature vector corresponding to the selected parameter D . This adaptive algorithm finds by itself the most 'distinguishing' points of an object.

A variation of this process is employed during the search of a query image, as follows. We convolve the query image with a set of Gabor functions. Assuming that we know the centre of the object in the query image, we form D_{max} feature vectors, one for each sampling geometry for $D=1..D_{max}$. The database itself has D_{max} sections and we search each section using the corresponding query feature vector for the closest match. If we do not know the centre of the object, which is usually the case, we apply this process many times, scanning the entire query image. Note, however, that during the search phase, only object features that were stored with a geometry identical to the one currently used by the search algorithm are compared.

This method can be applied to other problems where a meaningful representation of an object image is required with a small number of data points. Its advantages include automatic training capability, sparse object representation, low coding redundancy for the search algorithm, low storage requirements and increased search speeds. The disadvantages of this method are that it may not find all the distinguishing features and its relatively long training time.

4 Discrimination Merit Figure

In order to evaluate the quality of a particular set of sampling geometries for a given D_{max} , a merit figure calculation is called for. From an object recognition point of view, we are interested in finding a set of parameters that will maximize the inter-object separation or distance. The expression below presents the merit function, m , used for that purpose:

$$m = \sum_{l=1}^N \sum_{j=1}^{(N-1)/p} \frac{SD_{l,j}}{(N-1) \cdot \frac{1}{p}} \quad (6)$$

where SD_l is the sorted set of $N-1$ distance values, d_j between a target object l and the other objects in a library of size N as follows: $SD_l = \langle d_1, d_2, \dots, d_{N-1} \rangle$. We take into account the mean value of the p percent closest objects found at each iteration. In this work p was empirically set to 10 percent.

The reasoning behind this formula is simple. We want to make the distances between different objects in the database as large as possible. Therefore, we sort them and try to maximize the average of the p percent smallest distances. If the smallest distances are large enough, then we have achieved our goal.

5 Iterative Optimization

The work presented in this paper employs an adaptive sampling method for object representation. This method is required to determine the individual optimal sampling configurations to be applied to each target item in a database of object images. Assuming exhaustive search is used, the resulting number of possible permutations would be as large as $P_n = D_{max}^N$, where N is the number of targets in the database, and D_{max} is the number of possible sampling geometries. Even for a modest database containing 100 targets and a relatively small D_{max} value of four, there are 4^{100} possible permutations or roughly 10^{60} . Obviously, any exhaustive search method would perform poorly here in terms of computation times.

Figure 4 presents the pseudo code for an iterative optimization algorithm designed to alleviate this problem. The purpose of this algorithm is to determine, for each target in the database, the optimal sampling geometry that will maximize its distance function from all the other objects.

We begin by assigning, according to a uniform distribution, a random value for the sampling geometry D , to each object in the database. The iterative process finds, for each image, the configuration that will maximize a

feature-distance based merit function. A cumulative figure is computed for the entire iteration, which consists of a summation of the individual maximal merit figures found for each object.

T: Data structure containing pre-computed Gabor features for each target and for $D=1,2,\dots,D_{max}$
L: Array used for storage of initial, intermediate and final values of *D* for each target object.

```

IterativeOptimization(L,T)
Initialize L with uniformly distributed random values [1..Dmax]
WHILE(GLOBALprevious_merit != GLOBALcurrent_merit) DO
  GLOBALprevious_merit := GLOBALcurrent_merit;
  FOR i := 1 to N
    maxYield := MAXFLOAT;
    FOR d := 1 to Dmax
      FOR j := 1 to N
        m := CalculateLocalYield(d,i,j);
      END
      if (maxYield > m) maxYield := m; Max_d := d; //hold maximum local yield found so far
    END
    L(i) := Max_d; // store D value corresponding to maximum local yield in L
    GLOBALcurrent_merit := GLOBALcurrent_merit + maxYield;
  END
END WHILE
END IterativeOptimization.

```

Figure 4: Iterative Optimization Algorithm

The iterative process continues until there is no difference between successive cumulative merit figures. Experimental results show that for a D_{max} value of 20 and with 100 images the algorithm converged within less than 6 iterations. Further, our experience indicates that the final number of iterations is affected mainly by the size of the database and to a much lesser extent by the value of D_{max} .

5.1 Computational Complexity

Experimental results have shown that the number of iterations required for the algorithm to converge is in an order of magnitude smaller than the number of objects. Therefore, we can state with a degree of certainty that the proposed iterative optimization algorithm has quadratic complexity in the size of the object database given by $\Theta(N^2 D_{max} I)$, where N is the number of objects in the database, I is the number of iterations necessary for the algorithm to converge and D_{max} is the number of possible sampling geometries. For clarity purposes the similarity measure computation is not included in this calculation.

This complexity is significantly smaller than the one required for an optimal exhaustive search algorithm. For the case where $D_{max} = 4$ and $N=100$ the complexity of the iterative algorithm with $I = 6$ evaluates to $24 \cdot 10^4$, compared to 10^{60} operations required by a theoretical exhaustive search algorithm.

The number N of items in an object database can differ considerably. Industrial or manufacturing applications may not require more than a few hundred items in such a database. Conversely, a face database for law-enforcement purposes may have tens of thousands of entries. Thus, the learning time required for the proposed adaptive algorithm may substantially vary between applications.

5.2 Implementation Issues

For the experiments described in this work, the running time of the iterative optimization algorithm, for 100 base objects and with $D_{max} = 4$, was less than 78 seconds on a Sun SPARC Ultra-60. Note that the proposed algorithm was implemented in a multimedia interpreter environment (MediaMath), porting it would decrease execution times further.

An issue of importance is adding new objects to the database. For example, a database of faces used for law-enforcement purposes may have to be updated on a daily basis. Object databases are typically very large, thus, running time of the iterative optimization algorithm can be of concern.

A solution to this problem would be to keep, at all times, two copies of the raw data. One copy will be used by the running object recognition application while the other will have the new items appended to it. The latter copy will be used by the iterative optimization algorithm. At regular intervals e.g., nightly, the running copy will be exchanged by the newly created one, containing data of the new entries.

6 Experimental Results

This section presents the results of several experiments that were designed to demonstrate the effectiveness of the proposed adaptive algorithm. For the experiments described in this section we have created our own database of objects. To generate the object database, photographs were taken, using a digital camera with resolution 640 X 480, of the following items:

- 26 2D lower case latin alphabet letters, using Times New Roman font of average height equal to 5 centimeters or 50 pixels.
- 26 2D upper case latin alphabet letters, using Times New Roman font of average height equal to 7 centimeters or 70 pixels.
- 5 2D graphical symbols of average height equal to 7 centimeters or 70 pixels, each at three different orientations.
- 11 3D hand tools, each at three different orientations.

Quite a few of these items are similar in appearance. Each object was processed at 5 different scales resulting in a database of 500 items.

Figure 5 illustrates the dependency between the merit factor and the sampling-geometry as expressed by the value of D_{max} . The data for this figure was obtained by running the optimization algorithm on a down-sampled version of the targets in the object library. The scale factor used here (20%) determines the upper bound for the value of D_{max} . A too large value for D_{max} will result in samples crossing over the boundaries of the down-sampled target image.

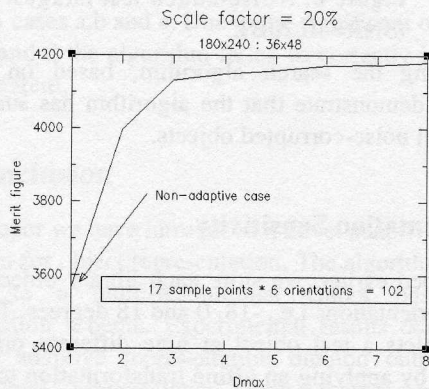


Figure 5: Merit figure change with D_{max} for scaled target objects.

The non-adaptive case corresponds to the condition where $D_{max}=1$. The merit figure monotonically increases with D_{max} . However, the rate of increase decreases. This can be attributed to the fact that most of the distinguishing features of target objects are located close to the centre of the target image. As we include in our feature set sample points that are further apart from the centre of a target image, their contribution to differentiating between similar images diminishes.

Because target images are scaled down, more distinguishing information is concentrated in a smaller image area. As a result, the merit figure curve saturates at or above a D_{max} value of 5. Increasing the area covered by the sampling geometry may not yield any significant additional feature separation.

Figure 6 illustrates the merit figure as a function of D_{max} for unscaled target objects. Higher D_{max} values are present here because the object images are larger and there is a smaller danger of crossovers. The merit-figure function has a higher rate of increase for the lower values of D_{max} . Conversely, the merit figure increases very little for D_{max} values exceeding 12 and asymptotically converges further on.

We generated several query images for the purpose of evaluating the presented algorithm. The query images were created separately from the images used for training and using hand alignment only. These images pose a difficult

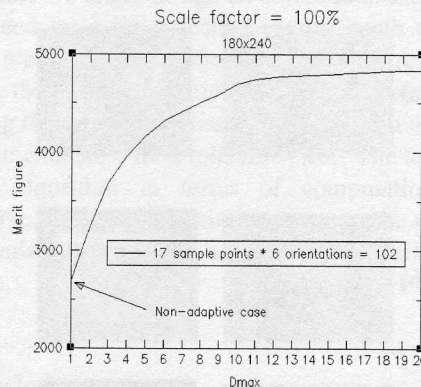


Figure 6: Merit figure Vs. D_{max} for unscaled target objects.

object recognition problem because they contain three or more different objects. In addition, all objects are closely grouped on top of a complex background containing fine texture. The experiments described in this section were carried out, for comparison purposes, with both an adaptive and non-adaptive representation algorithms. As mentioned earlier, the non-adaptive case can be viewed as a special case of adaptive sampling with $D_{max}=1$.

The search algorithm has two phases. Prior to the first phase, the query image is convolved with six Gabor filters at 30 degrees orientation increments. We have selected a central frequency of 1.32 radians/pixel and a standard deviation of 2.6 pixels. During the initial phase each pixel is assigned a label corresponding to the closest matching object in the library, together with a value indicating the distance from that object. We use Euclidean distance because of its robustness and low computational costs. The second phase of the algorithm employs statistical methods to produce a list of candidate objects in order of descending likelihood of presence.

The adaptive-sampling based search algorithm has successfully recognized all the objects in the test images presented in Figure 7. We have used a dynamic sampling geometry with $F=2$ and 17 sample points for a feature vector length of 102. However, a non-adaptive sampling based search algorithm produced significantly inferior results. Only between 25% (Figure 7 d) and 75% (Figure 7 c) of the objects in the query images were correctly detected. In addition, the length of the detected object list, which also includes false entries, was longer by an average of 50% when non-adaptive sampling was used. A longer object list increases the difficulty of identifying false matches.

By substituting the pixels in the original image with their corresponding closest-matching-object distance value, we obtain the images in the right column of Figure 7. Darker areas, corresponding to lower distances, are

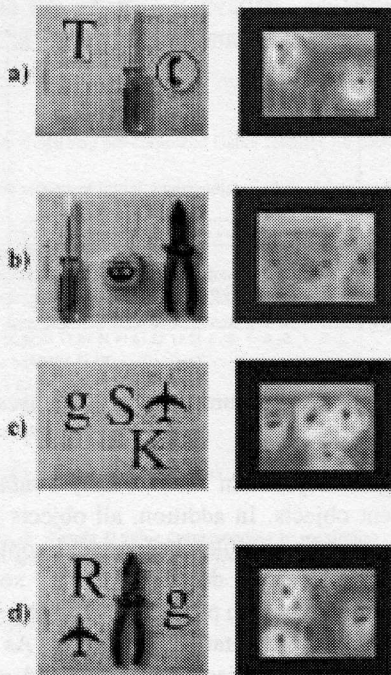


Figure 7: Sample query images.

generated around the centres of objects present in the query images. This information can be used to identify the position of selected items in searched images.

Because of the sampling geometry, a narrow band is removed from the sides of the image resulting in the black margins depicted. In the scaled image, that band is no wider than half of the maximum width of the sampling-geometry used, typically a few pixels.

In order to obtain recognition results of similar quality, a non-adaptive representation scheme required a sampling geometry with 33 sample points for a feature vector of 198 elements. This is a difference of 48% compared with adaptive representation.

6.1 Noise Sensitivity

One of the most significant strengths of the adaptive sampling method presented in the work is its increased noise tolerance. By maximizing the inter-object distance, the probability of incorrect classification due to noise is minimized. In order to evaluate the robustness of the proposed algorithm, we added Gaussian noise at a Signal to Noise Ratio (SNR) of -0.2 dB. Images containing gray-scale and binary objects were selected for detailed study as follows.

The above figure depicts four examples of images containing a single object with added noise. The noise is added to the grey-level values of each pixel and severely degrades the images under examination. Results obtained

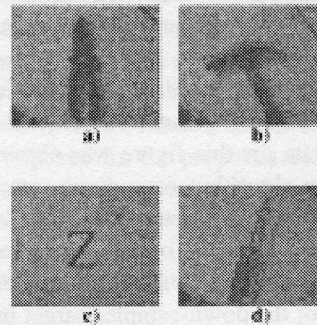


Figure 8: Noise-added test images (SNR= -0.2dB).

by applying the search algorithm, based on adaptive sampling, demonstrate that the algorithm has successfully detected all noise-corrupted objects.

6.2 Orientation Sensitivity

As described earlier, we store each library object at three different orientations i.e., -18, 0 and 18 degrees. The figure below depicts a test object at nine different orientations generated by applying an affine transformation to the base (0°) object image. The angular spacing is equal to 8° with orientation values ranging from -32° to 32° .

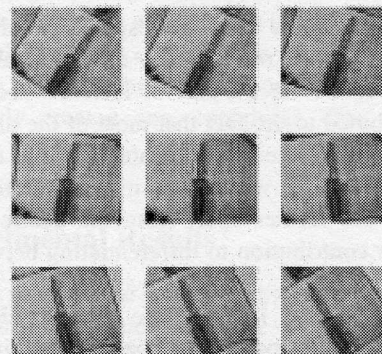


Figure 9: Rotation examples.

Although each object is stored in the object library with only two orientations of ± 18 degrees, the algorithm is able to correctly identify objects with rotation up to ± 32 degrees or an 88% angular increase.

6.3 Occlusion Sensitivity

Occlusion can occur in several real-life situations where an object in an image is partially covered by another object or is subjected to shading.

The figure above depicts five different occlusion scenarios. A vertical dark bar is sled across the test object, from left to right. The vertical and horizontal dimensions of the occluding bar are 170 pixels and 20 pixels, respectively.

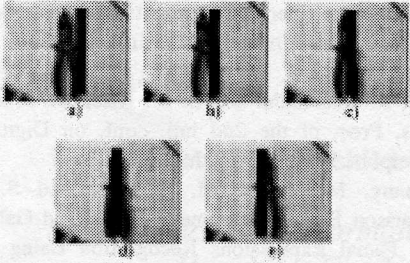


Figure 10: Occlusion examples.

The proposed algorithm has successfully detected the object in cases a,b and e. Due to the significant occlusion in cases c and d the algorithm failed to correctly identify the occluded item.

7 Conclusion

In this paper we have introduced a novel adaptive-sampling algorithm for object representation. The algorithm has been applied to a multi-resolucional Gabor based object identification scheme. Experimental results demonstrated that the adaptive representation method outperforms a conventional representation with feature-vector sizes longer by up to 48%. The properties of the proposed algorithm can be utilized for dimensionality reduction of feature vectors in Gabor and other transform spaces.

A merit function expressing the obtained relative degree of object discrimination has been devised. We have showed that the merit function is monotonically increasing with the sampling geometry maximum allowed expansion parameter. This proves that the separation between objects, in terms of the distance metric of choice, increases as additional allowed permutations of the sampling geometry are introduced. However, an upper bound exists for the yield function beyond which increasing the span of the sampling geometry will not result in further improvement in the merit function value. This phenomena is due to the fact that the sampling geometry at high span values also covers areas that are outside the boundaries belonging to the object of interest.

In order to reduce computational demands, iterative optimization was applied. This method significantly decreases the computational requirements of the algorithm which would be otherwise prohibitive. The computational complexity of the training phase of the algorithm depends upon the number of objects in the database, the maximal degree of freedom of the sampling geometry and on the number of iterations required by the optimization to converge. In practice, the number of iterations is correlated with the number of objects and does not exceed a reasonably small fraction of it.

Assume we are to design an object database while being constrained by a fixed feature-vector length requirement. Clearly we would opt, regardless of other restrictions, to employ the adaptive-sampling method described here. Applying this method increases the recognition quality of a system in exchange for a tolerable price. The learning costs of this algorithm in terms of computations are not prohibitive. Further, the training phase can be done off line to minimize the interference with the operation of the database.

Acknowledgment

The support of the NSERC (App. No. OGPOO46645) is gratefully acknowledged.

References

- [1] A. D. Alexandrov, W. Y. Ma, A. El Abbadi, and B. S. Manjunath, Adaptive Filtering and Indexing for Image Databases, Proc. SPIE Conference on Image and Video Databases, Vol. SPIE-2420, pp. 12-23, 1995.
- [2] J. Ben-Arie and W. Zhiqian, Pictorial recognition using affine-invariant spectral signatures, Proceedings, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 34-39, 1997.
- [3] J. Ben-Arie and P. Pandit P, A comparison of Gabor projection and decomposition methods in object detection, IEE Electronic Letters, Vol. 36, No. 21, pp. 1764-1766, 2000.
- [4] M. Betke and N. C. Makris, Fast object recognition in noisy images using simulated annealing. In Fifth International Conference on Computer Vision (Cambridge, MA, June 20-23, 1995), Vol. 1, pp. 523-530, Washington, DC., 1995.
- [5] C. Chen and D. Chen, Multi-resolucional Gabor filter in texture analysis, Pattern Recognition Letters, Vol. 17, No. 10, pp. 1069-1076, 1996.
- [6] K. Chung, S. Kee and S. Kim, Face recognition using principal component analysis of Gabor filter responses, Proceedings, International Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, pp. 53-57, 1999.
- [7] M. Dailey and G. Cottrell, PCA = Gabor for Expression Recognition, UCSD Computer Science and Engineering Technical Report CS-629, August 1999.
- [8] Y. Hamamoto, S. Uchimura, M. Watanabe, T. Yasuda, Y. Mitani and S. Tomoita, A Gabor Filter-Based Method for Recognizing Handwritten Numerals, Pattern Recognition, Vol. 31, No. 4, pp. 395-400, 1998.
- [9] A.K. Jain and S. Bhattacharjee, Address Block Location on Envelopes Using Gabor Filters, Proc. 11th Int'l. Conference on Pattern Recognition, The Hague, pp. 264-267, August 1992.

- [10] A. Jain, N. Ratha and S. Lakshmanan, Object Detection using Gabor Filters, Pattern Recognition, Vol. 30, No. 2, pp. 295-309, 1997.
- [11] W. Y. Ma and B. S. Manjunath, A texture thesaurus for browsing large aerial photographs, Journal of the American Society for Information Science, Vol. 49, No. 7, pp. 633-648, 1998.
- [12] B. Schiele and A. Pentland, Probabilistic object recognition and localization, Proceedings of the Seventh IEEE International Conference on Computer Vision, Vol. 1, pp. 177-182, 1999.
- [13] L. Wiskott, J. Fellous, N. Kruger and C. von der Malsburg, Face Recognition by Elastic Bunch Graph Matching, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, No. 7, pp. 775-779, 1997.
- [14] T. P. Weldon and W. E. Higgins, An algorithm for designing multiple Gabor filters for segmenting multi-textured images, IEEE Intl. Conf. Image Processing, Chicago, IL, pp. 4-7, October 1998.
- [15] V. Wu, R. Manmatha and E. M. Riseman, Finding Text In Images, Proc. of the 2nd intl. conf. on Digital Libraries, Philadelphia, PA, pp. 1-10, July 1997.
- [16] Z. Zhang, M. Lyons, M. Schuster and S. Akamatsu, Comparison Between Geometry-Based and Gabor-Wavelet-Based Facial Expressions Recognition Using Multi-Layer Perceptron. Proceedings, Third IEEE International Conference on Automatic Face and Gesture Recognition, IEEE Computer Society, pp. 454-459, 1998.