Template basis techniques to Pattern Recognition

Mark E. Lehr and Keh-Shin Lii

University of California Riverside, Department of Statistics
Riverside, California 92521

ABSTRACT

The purpose of this paper is to present an approach to pattern recognition which acknowledges theories in the fields of perception, the human visual system and decomposition. It is hoped that by taking a panoramic view of the subject matter that insights into the issue may uncover a possible course of action. At the very least, there should be evidence that certain tools used in the process adequately fit the analysis of the problem.

The contemporary perceptual theory to classification recognizes the fundamental concepts of scale and localization. The visual system can solve problems that would be intractable using a single depiction of a scene by having access to representations at different spatial scales. Enhancement, analysis and compression are the areas of image processing most germane to pattern recognition. And the simplest statistical approach to identifying patterns utilizes templates.

All these properties can be inherently exploited in the wavelet domain. By unifying the process of noise reduction, segmentation, feature extraction and classification it is possible to develop a general technique which might not be optimal but has the advantage of being computationally efficient. All the desired properties that are required in an analytic task of this nature are captured with multiresolution analysis.

Keywords: classification, detection, false alarm, feature extraction, matched filter, pattern recognition, perception, scale, segmentation.

1. RECOGNITION PRELIMINARIES

1.1 The Shape of Things

What defines an object and its form? Our ability to perceive or mentally grasp these concepts seems trivial since all living things with sight have this innate ability but they are nonetheless not easily quantifiable. Various schools of thought have tried to answer these questions and significant headway has been made by understanding the physiological as well as the psychological process of vision.

It is relevant at this point to determine the basis for the theories involved with biological characterization of the environment. This will give possible directions to take when pattern comparisons are attempted and how they might relate to using wavelets as part of the algorithmic process. The first area of consideration is perception which is distinguished by a signals detection, discrimination, and identification.[1]

Detection involves the transduction of signals from one form to another. This must not only be accomplished in a noisy environment but the transducer itself can limit or

0-8194-2213-4/96/$6.00

enhance the detection process. The visual system involves much more than just the eye. Once the light has been focused on the retina, many layers of processing must occur which emphasizes or rejects information thereby shaping the signal into something which can be interpreted.

This interpretation involves the ability to discriminate and identify. Is the object of interest a part of the scene and specifically which object out of many is of particular interest? It is a fact that these processes are accomplished in the visual system quite effectively with biological adaptation. The process of detection has been studied and significantly understood. The discrimination and identification processes occur rapidly and accurately with little thought as to how. This is exactly the problem, i.e. how to define a procedure or algorithm which mimics these two processes. They occur at a level that can belie significant understanding.

There have been three schools of thought which have endeavored to answer the question regarding the essence of form and shape. Naturally when theories fail in some regards later schools of thought pick what works and discard the rest. The first school is termed the structuralist concept of perception. This was popular at the turn of this century. Its main idea is simple and thereby attractive in that complex mental processes are created by combining fundamental components. Edward Titchener championed this method but flaws developed with the approach because it failed to quantify the unique elementary units.[1]

The next school of thought emphasized overall structure or pattern as the major determinant of form i.e. Gestalt which is German for form. Wertheimer, Kohler, and Koffka were expositors of this theory at the beginning of this century.[1] Principles which define this organization include, proximity, similarity, closure and continuation. While discovering important principles of perceptual organization they were unable to develop a lasting theory of the neural origins of these principles.

And finally, the contemporary approach which incorporates the best of both by recognizing the concept of scale. A scene is composed of many different building blocks which has information at many different spatial scales. By having access to representations at different spatial scales, the visual system can solve problems that would be intractable using a single representation.[2] In a natural image, intensity changes occur over a wide range of scales which are detected separately. Intensity changes in images arise from surface discontinuities, or reflectance, or illumination boundaries, and these all have the property that they are spatially localized.[3] The brain has the ability to separate form recognition from size regarding the invariance of the central presentation of visual stimuli. Thus, the mechanism of invariance to changes in size has been identified and localized as a function occurring in the left hemisphere.[4]

Therefore, the most up to date information on the ability of human beings to perceive and comprehend images relies on the concept of scale and spatial localization.

An analysis tool with these characteristics would have the advantage of matching the human visual systems abilities.

1.2 Template Classification

The term pattern recognition refers simply to the process of identifying an unknown object in an image. The term pattern refers to a collection of features derived from an object to be identified.

In general there are three basic approaches to pattern classification: statistical, syntactic, and neural methods.[9] The first relies on defining a set of decision rules based on standard statistical theory. A set of characteristic feature measurements are extracted from the input image which are then combined to define a feature vector. The decision rules for this assignment are derived either from a priori knowledge of the expected distribution of the pattern classes or from knowledge acquired through a training process involving many initial measurements.

Template matching is the simplest statistical classification approach. The template matching process involves moving the template to every position in the image and evaluating the degree of similarity at each position. This also explains why the process of template matching is sometimes described as matched filtering.[10] A computational modification can be made to reduce the search process by segmenting the image. This reduces the search area by identifying possible objects to be matched. A regression then is necessary to find the parameters associated with scale and rotation.

1.3 Image Processing Operations

Image processing operations can be decomposed into five classes.[5] Nonetheless, any given image may require one or more operations dependent upon the application. Pattern recognition techniques exploit the areas of enhancement, analysis and compression.

The first class of image processing is enhancement. This operation might improve brightness and contrast characteristics, sharpen and clarify details as well as reduce various types of noise. Wavelets could certainly be utilized in this type of operation as suggested by wavelet shrinkage methods.[6]

The second class is that of image restoration. The images that require restoration are normally associated with problems in the imaging system. Given that this problem can be characterized then it is sometimes possible to remove effects such as geometric distortion, camera motion or even improper focus i.e. the Hubble space telescope.

The next class is one of the most important and most often used which is that of image analysis. Analysis quantifies elements in an image, such as size, shape, and other statistical attributes. The desired goal is to measure and classify image information which is necessary in identifying objects. For pattern recognition one of the most important operations is that of finding objects in an image. Filters are normally used to segment an image. A major reduction in the search space can be achieved by segmentation--the division of the image into sets of related features. Without segmentation, a model would have to be matched against all possible combinations of features in the image, so good segmentation is crucial for reducing the combinatorics of this search. Segmentation has long been recognized as a central problem and most methods have been based on region analysis or scene-specific measures rather than on general methods of perceptual organization.[7]

Image compression operations reduce the data content necessary to describe an image. This is possible since most images contain large amounts of redundant information. From a pattern classification perspective, compression simply results in less to compare [8]

The last operation is associated with synthesizing information. One form of this operation involves the reconstruction of an image using multiple projections. The other is merely a visualization technique for representing data. The intent of these operations is to create images that convey important features found in the data that might otherwise be very difficult or impossible to detect.

Therefore, classifying an image would best utilize the processing elements of enhancement, segmentation and reduction in the parameter space which are all image processing techniques that apply to wavelet domain analysis.

2. THE WAVELET ADVANTAGE

There are many forms of efficiency possible by exploiting wavelets in a template approach. The tool is ultimately suited to this type of investigation since we obtain at least six advantages that are inherent in this analysis.

1. Scale considerations,
2. Localization properties,
3. Noise reduction,
4. Segmentation with orientation preference,
5. Feature extraction/reduction, and
6. Computational efficiency

Advantages 1 and 2 correspond to the human visual system and the modern approach to perception. Advantages 3 through 5 to the operations normally associated with image processing. The last advantage corresponds to the implementation of pyramidal algorithms which can efficiently compute various types of statistical descriptions (weighted averages, moments, etc.) of the data in the image blocks, and can efficiently perform coarse operations on an image by combining these descriptions.[11]

3. METHODOLOGY

A template is a pattern or guide with which to compare to an object. A proposed approach which minimizes computational effort would utilize a tool matched to the problem at hand. Therefore, a library of templates are decomposed into their wavelet coefficients. The template or matched filter in this case are the dictionary of patterns that have been decomposed into the wavelet domain.

A template T in the picture domain can be represented by the sum of two-dimensional wavelets given by

$$T(x,y) \approx \sum_{m,n} s_{J,m,n}\phi_{J,m,n}(x,y) + \sum_{o\in\{h,d,v\}} \sum_{j=1}^{J} \sum_{m,n} d^{o}_{j,m,n}\psi^{o}_{j,m,n}(x,y). \tag{1}$$

A template W_T in the wavelet domain is portrayed by a partitioned matrix of the coefficients s and d from equation (1) illustrated by

$$W_T = \left[\begin{array}{c|c} d^h & d^d \\ \hline s & d^v \end{array}\right] \tag{2}$$

where s coefficients represent the smoothed low pass filter output associated with the father wavelet ϕ at scale J and translation m and n. The d coefficients represent the detail at scale j associated with the high pass filter output of the mother wavelet ψ. The orientation of the filters corresponds to horizontal d^h, vertical d^v and diagonal d^d segmentation filters.[12] With proper scaling and placement in the resulting matrix the coefficients resemble a scaled version of the original image having primal sketch characteristics.[3]

The image I to be analyzed is represented similarly with the same analyzing wavelet by

$$I(x,y) \approx \sum_{m',n'} s'_{J',m',n'}\phi_{J',m',n'}(x,y) + \sum_{o\in\{h,d,v\}} \sum_{j'=1}^{J'} \sum_{m',n'} d'^{o}_{j',m',n'}\psi^{o}_{j',m',n'}(x,y). \tag{3}$$

This is represented by W_I in the wavelet domain with

$$W_I = \left[\begin{array}{c|c} d'^h & d'^d \\ \hline s' & d'^v \end{array}\right]. \tag{4}$$

Object O refers to the segmentation of the image using various coefficients of the mother wavelets at the lowest levels

$$O(x,y) \approx g(d^{\prime o}_{j',m',n'}). \tag{5}$$

The d coefficients then comprise the elements utilized in the distance measurement. A preliminary step would take some statistics with respect to the object identifying high level feature characteristics to eliminate those templates that have a high degree of divergence with the object.. The distance measurement D using some metric is given by

$$D = metric\left|\underline{c}(f,tx,ty) - \underline{c'}\right| \tag{6}$$

where the c and c' vectors are composed of elements from the W_T and W_I matrix. The elements of the template in this case will need to be scaled by f, translated and rotated by tx and ty before the metric is determined. This can be obtained with a least squares approach and aligned using an interpolation procedure. In practice, the vector which is to be scaled and rotated could either be the object or the template, but it is felt that the template has not been degraded in the imaging process and thus is the better candidate. This task is not trivial and will require a significant amount of thought to optimize the outcome.

In practice the peak signal to noise ratio (PSNR) is used to measure the difference between two images before and after compression or when noise is added to the image. If we choose the metric as the root mean square (rms) difference of the object and the template, then

$$PSNR = 20\log_{10}\left(\frac{b}{D}\right) \tag{7}$$

where the maximum signal b can be ascertained from the image whereas the rms error can be determined simultaneously with equation (6) if the appropriate metric is chosen to identify the discriminant.

The distance measurement D will be used to identify an object by comparing to a threshold. If a number of templates match the object of interest then more coefficients in the wavelet domain will be used to break the tie or the distance measurement with the smallest metric will be chosen. Threshold ε is set to give a probability of detection (P_D) and false alarm (P_{FA}) based on the PSNR as defined by

$$P_D = \Pr(D \le \varepsilon | O_k = T_l) = h(O_k, T_l, \varepsilon, PSNR), and \tag{8}$$

$$P_{FA} = \Pr(D \le \varepsilon | O_k \ne T_l) = h(O_k, T_l, \varepsilon, PSNR). \tag{9}$$

Since the noise can be estimated from the wavelet coefficients, then the threshold levels can be stored prior to the analysis. The number of coefficients to be stored and compared is ultimately based on the discrimination required between template classes given the peak signal to noise ratio and the energy compaction of the wavelet analyzing function.[13]

4. A SIMPLE ILLUSTRATION

Figure 1 gives a simple illustration of the proposed process. Panel (a) contains a template of an eye. Panel (b) is the Haar wavelet transform of the eye which is represented by equations (1) and (2) with J=1. Similarly panels (c) and (d) contain the image and its wavelet transform represented in equations (3) and (4). The object is best identified by the horizontal coefficients d^h in the upper left partition. A simple rms metric is used from equation (6) which has been scaled but not rotated. Since no noise has been added to the image there is no need to set a threshold level with the PSNR from equation (7). The probability of detection and false alarm are 100% and 0% utilizing a threshold level of zero by equation (8) and (9). A match was made with the horizontal partition but the other 3 positions would contain the classification as well as depicted in panel (e). These directly translate to the position in the image domain contained in panel (f).

Figure 2 gives an illustration of the same process with noise. Panels (a) and (b) contain the image and its wavelet transform represented in equations (3) and (4). The noise level from equation (7) is 20db. The threshold has been set at the noise level estimated by the d coefficients. Panel (c) contains the resulting match of the upper left partition d^h coefficients. The P_{FA} is 21% of the surface area using the single partition. Panel (d) contains the resulting match of the upper two partitions d^h and d^d. The P_{FA} is 9% of the surface area using these two partitions. And finally, panel (e) contains the match using all d partitions and the template is correctly identified. Panel (f) replaces the template of the eye in the noisy image.

5. DISCUSSION

A number of areas regarding the actual implementation of the process described above need further investigation. For instance, the number of coefficients and thus the feature vector elements would be reduced should the template and wavelet analyzing function match thereby decreasing the comparison time. However, a different analyzing wavelet would have to be used for each class of templates and the savings in computations for the comparison might not be worth the effort of using multiple wavelets.

At each level there are three different detail matrix coefficients d^h, d^d, and d^v to compare. Which orientation and at what level would best be used in the comparison? This may not be of importance since when one is found all the rest are known due to the redundancy. The example with additive noise shows that using all three significantly reduces the false alarm rate and the comparisons only need to be made with those that match in previous partitions.

The scale and orientation refinements along with the interpolation process are however, prime areas for theoretical investigation.[13,14]

Figure 1 No Noise Example (a)

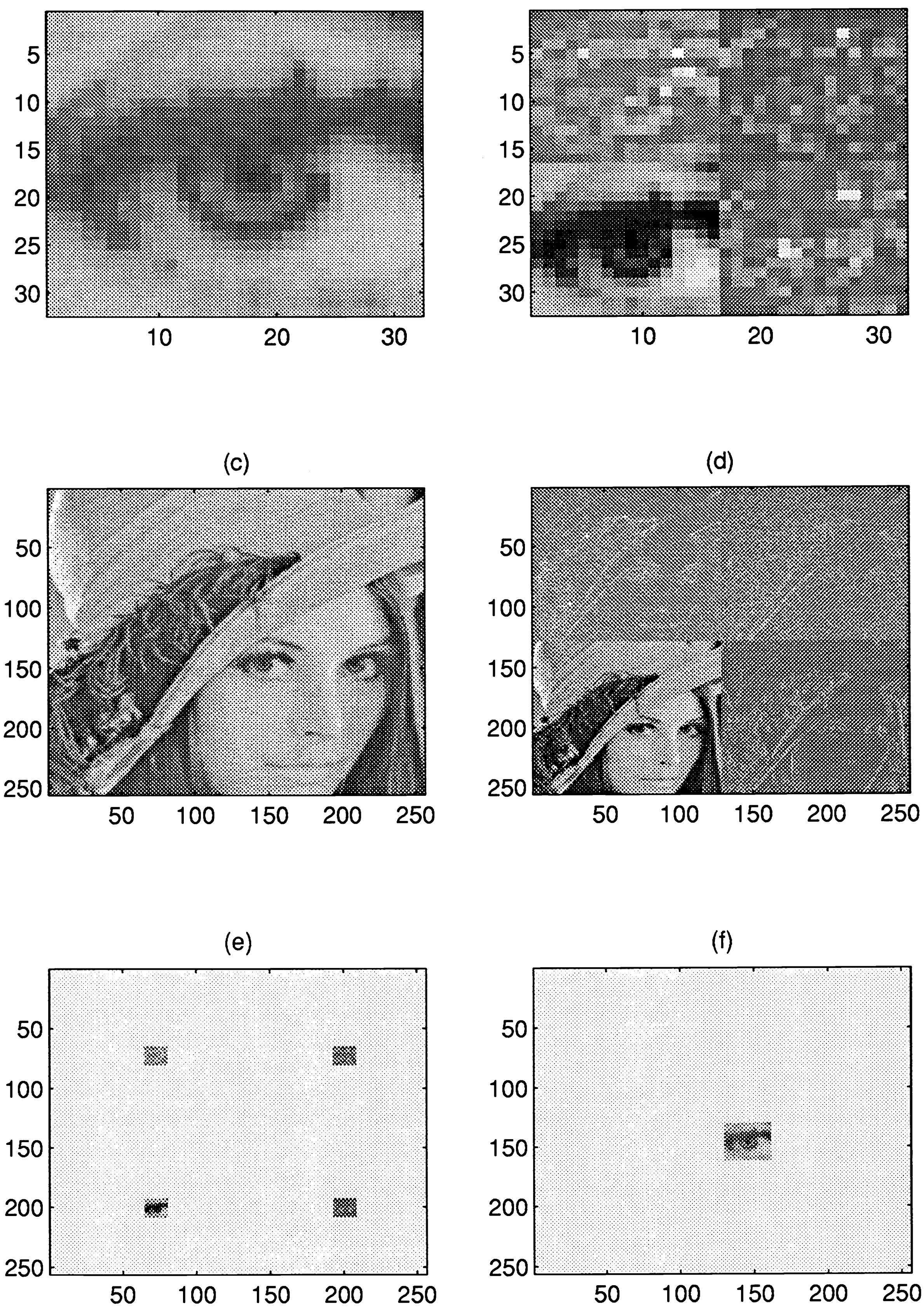
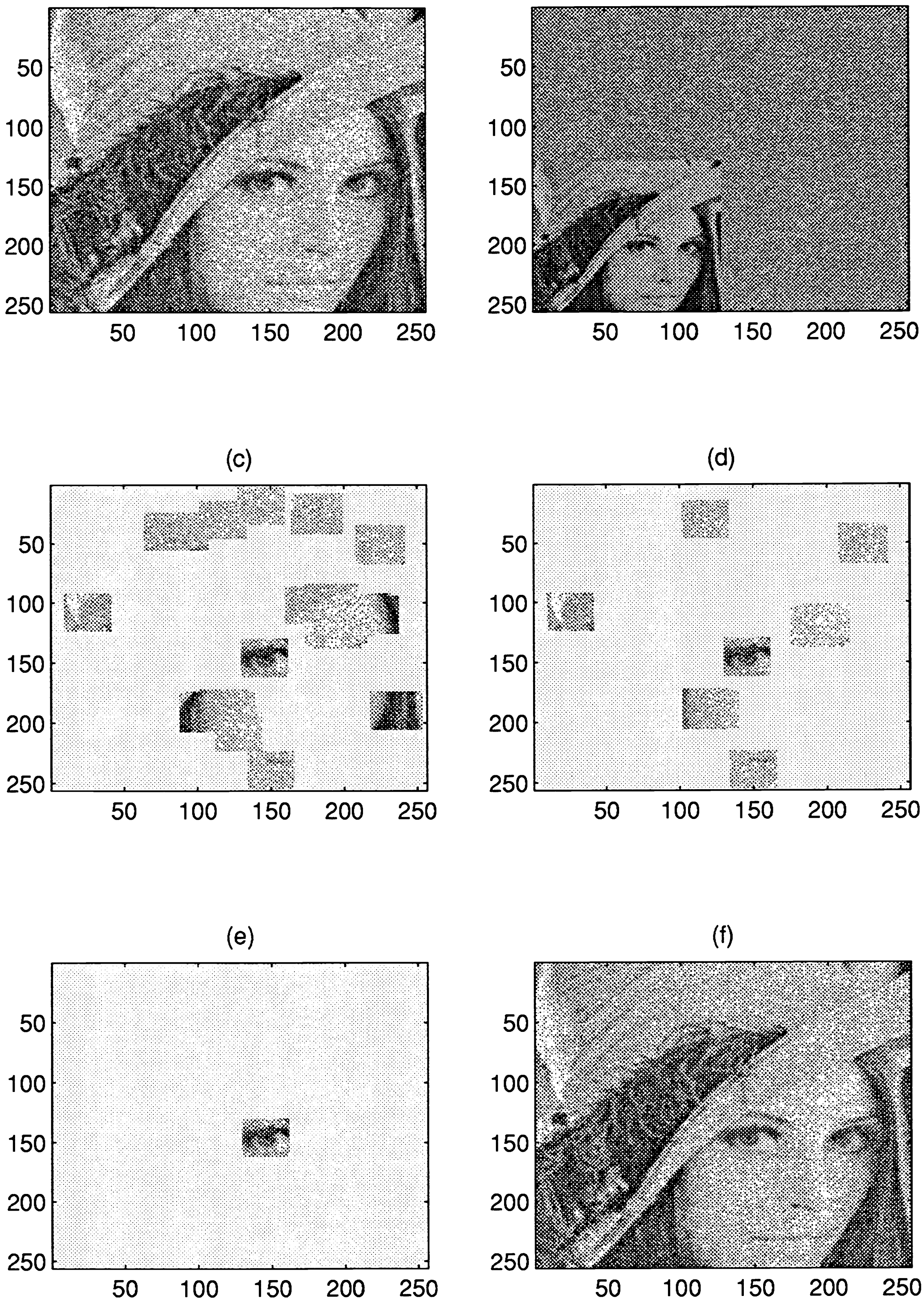


Figure 2 Noise Example (a)



6. ACKNOWLEDGMENTS

This research was supported by ONR N00014-92-J-1086 and N00014-93-1-0892.

7. REFERENCES

1. Robert Sekuler and Randolph Blake, *Perception,* Chapter 5, third edition, Mcgraw-Hill, New York , 1994.
2. D. Marr and E. Hildreth, "Theory of edge detection", *Proceedings Royal Society of London*, B, 207, 187-217, Great Britain, 1980.
3. David Marr, *Vision*, pp 44-50, W.H. Freeman and Company, San Francisco, 1982.
4. Vadim D. Glezer, *Vision and Mind*, pp 196, Lawrence Erlbaum Associates, Mahwah, New Jersey, 1995.
5. Gregory A. Baxes, *Digital Image Processing*, Chapters 4-7, John Wiley and Sons, New York, 1994.
6. David L. Donoho and Iain M. Johnstone and Gerard Kerkyacharian and Dominique Picard, "Wavelet Shrinkage: Asymtopia", *JRSS*, B, 57, 2, 1995.
7. David G. Lowe, *Perceptual Organization and Visual Recognition*, pp 5, 19, Kluwer Academic Publishers, Boston, 1985.
8. Luc Devroye and Laszlo Gyorfi and Gabor Lugosi, *A Probabilistic Theory of Pattern Recognition*, Chapter 32, Springer Verlag, New York, 1996.
9. Robert Schalkoff, *Pattern Recognition: Statistical, Structural, and Neural Approaches*, Chapters 2,6,10, John Wiley and Sons, 1992.
10. G. W. Awcock and R. Thomas, *Applied Image Processing*, Chapter 7, McGraw-Hill, New York, 1996.
11. A. Rosenfeld, "Pyramid algorithms for efficient vision", *Vision: coding and efficiency*, pp 412,-422, edited by Colin Blakemore, Cambridge University Press, Cambridge, 1990.
12. Andrew Bruce and Hong-Ye Gao, *S+ Wavelets*, Chapter 4, StatSci Division MathSoft Inc., Seattle, Washington, 1994.
13. William H. Press and Saul A. Teukolsky and William T. Vetterling and Brian P. Flannery, *Numerical Recipes in C*, pp 576-584,602-606, second edition, Cambridge University Press, Cambridge, 1992.
14. Yuval Fisher, *Fractal Image Compression*, pp 20-21, Springer Verlag, New York, 1995.