Fast Pattern Recognition Using Gradient-Descent Search in an Image Pyramid

James MacLean
University of Toronto
Department of Computer Science
Toronto, Canada, M5S 1A1
maclean@vis.toronto.edu

John Tsotsos
York University
Department of Computer Science
4700 Keele St., Toronto, Canada, M3J 1P3
tsotsos@cs.yorku.ca

Abstract

A new technique for fast pattern recognition using normalized grey-scale correlation (NGC) is described. While NGC has traditionally been slow due to computational intensity issues, the introduction of both a pyramid structure and a local estimate of the correlation surface gradient allows for recognition in 10–50 ms using modest microcomputer hardware. The algorithm is designed to analyze the target off-line prior to starting the search. Issues surrounding determining an appropriate depth for the pyramid representation and performing sub-pixel localization of the target instance are discussed. The speed and robustness of the method makes it attractive for industrial applications.

1. Introduction

Fast pattern recognition has long been of interest to industry. The ability to quickly locate a desired part, or accurately determine its location for registration purposes is a valuable tool in manufacturing. One popular method is *normalized grey-scale correlation* (NGC), in which a target image is correlated with the entire search region, and peak responses used to identify the location of the target. In the past specialized hardware was often required to do this with enough speed to make it viable, but recent advances in microcomputers have made it possible to perform searches on non-specialized hardware. This paper describes new techniques to allow for fast target locating using modest, non-specialized microcomputer hardware. More specifically, the algorithm is suited to implementation on a personal computer equipped with only an image acquisition board and a camera.

The method has two main features. The first is the use of an image pyramid representation for both the image to be searched and the target or pattern to be recognized. The second is the use of gradient descent search of the correlation surface to find a maximum match. A method for subpixel localization of the target instance has also been derived.

This paper will begin by briefly discussing some related previous work. It will then provide a description of the method, present results and discuss some of the strengths and weaknesses of the proposed technique. Further details may be found in [9].

2 Previous Work

Some of the earliest attempts at pattern recognition have been done using correlation methods, which are reviewed in detail in [4, 10, 2]. The concept of normalized correlation was developed to combat the effect of different illumination levels on correlation techniques. A major downfall of correlation techniques is that they are computationally very intensive, and as such tend to be slow. Another class of techniques involves matching moments [10], but is limited to the case where the pattern to be recognized can be easily segmented from the background prior to recognition. When this can be done these techniques are often powerful, as they can provide rotation and size invariant recognition [5, 3].

A new class of techniques are centred around neural networks, but to date they are more suited to pattern classification than pattern localization. Kulkarni [7] describes a size/rotation invariant object recognition method using back propagation to recognize feature vectors, but the vectors are based on moments and thus suffer from the segmentation problem mentioned above. Work done by Wechsler & Zimmerman [12, 13] looks more promising, but seems too complex to be fast on modest hardware.

The concept of using pyramid representations for image analysis is certainly not new, and is closely linked to the ideas of the emerging "scale-space" field [6, 8]. In this paper pyramid representations are used to overcome some of the computational problems associated with NGC. Burt [1] has constructed a pyramid-based attention model which uses Gaussian and Laplacian pyramids, and which is used for foveation, tracking and determining where to look next based on image saliency. Their system can be used for pattern recognition by using a "pattern tree" to represent salient features of objects at different resolutions. If enough features are matched, then the

pattern has been found. Special hardware has been developed [11] to aid in the construction of image pyramids.

3 Description of Method

The technique for locating target instances has several major components. The first is a pyramid representation used for both the target image and the image to be searched. The second is the representation used for the correlation gradient surface for the gradient descent search. These are used in a search framework which performs full NGC at the top level of the pyramid, and uses the results from this to guide the search for probable candidates using gradient descent in the pyramid representation. Finally, the results can be refined using a subpixel localization method.

3.1 Building the Image Pyramid

The first step in the search process is building a pyramid representation of both the target and the image. The pyramid is based on reducing the dimensions of the image by a factor of 2 at each level. Assume we start with an image $I(x, y)$ of dimension $I_h \times I_w$, and let $I^k(x, y)$ be the image at the kth level of the pyramid ($I^0 = I$). Each pixel in level k is the average value of 4 pixels at level $k-1$, *i.e.* $I^k(x, y) = [I^{k-1}(2x, 2y) + I^{k-1}(2x+1, 2y) + I^{k-1}(2x, 2y+1) + I^{k-1}(2x+1, 2y+1)]/4$. The pyramid can be built quickly since each pixel is computed using 3 adds and 1 shift operation, and the entire pyramid fits into less than twice the memory of the original image. The number of levels in the pyramid is limited to $k_{\max} \leq \log_2 \min(I_w, I_h))$. An example of a pyramid with 3 levels is given in Figure 1. Our pyramid representation is simpler than that used in [1] in order that it may be implemented without special hardware.

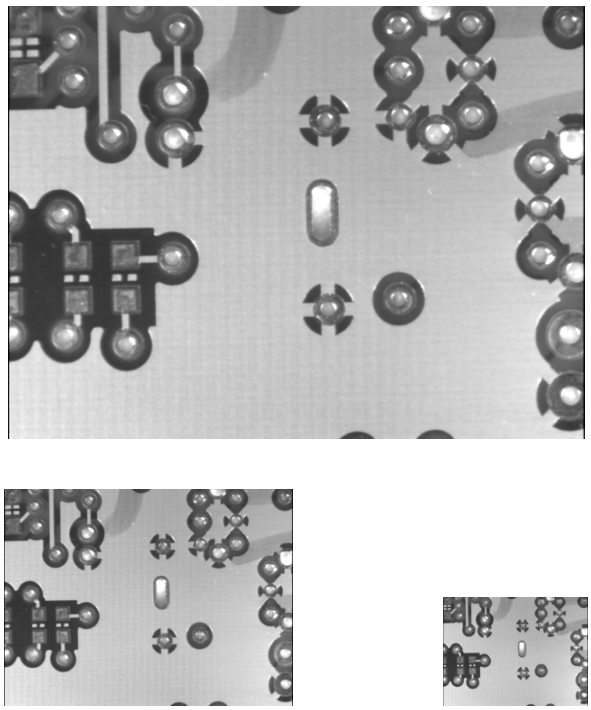

Figure 1. The pyramid representation for a typical image is shown. The pyramid has three levels, with level 0 being the largest image (top), and level 2 being the smallest (bottom right). In the level 0 image, a search target is defined.

3.1.1 Tuning The Pyramid

The question arises, "how many levels should be used in the target pyramid?" While a maximum limit is given in the previous section, the practical limit is typically much smaller. For example, consider the following pathological case: the target consists of an alternating checkerboard pattern of black and white pixels. After one level of the pyramid, all that is left is uniform grey! Another source of match degradation occurs due to our use of a non-overlapping pyramid representation. Slight shifts of the target instance in the base image may lead to different representations at higher levels of the pyramid [9]. Therefore the depth of the pyramid must be chosen according to the characteristics of the target itself.

A "worst case" analysis technique is used. Pyramids of maximum height are constructed for the target and shifted versions of the target. The target is then correlated against the shifted versions of itself at each level to determine worst possible scores. Maximum pyramid depth is then determined by

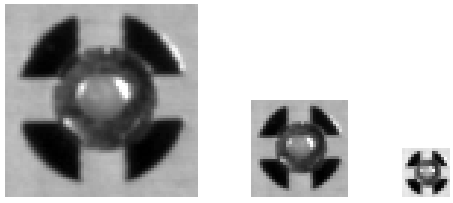

Figure 2. The pyramid representation of the target in Figure 1 is shown. The pyramid has the same number of levels as the image pyramid.

choosing the largest value for $k_{\max}$ such that the worst score at that level is still considered "acceptable". Acceptable scores can be quite low, as we really only need to identify possible match candidates. Each candidate is then verified at all levels of the pyramid before declaring it a true match.

3.2 Derivation of Correlation Gradient

Instead of doing a full correlation at each level of the pyramid, it is possible to compute a local estimate of the gradient of the correlation surface. This has the advantage that we can compute only the gradient points we need, since the value of the gradient itself tells us where to look next.

Assuming that $M(x, y)$ is a target image of size $M_w \times M_h$, and that $W(x, y)$ is a windowing function of the same size whose gradient goes to zero at the edges, then the (continuous) normalized correlation is given by

$$C(u, v) = \frac{\iint I(x, y)(MW)(x - u, y - v)dx\, dy}{\left[\iint I^2(x, y)W(x - u, y - v)dx\, dy\right]^{1/2}} . \quad (1)$$

It is assumed that M is normalized with respect to W, *i.e.* that $\iint (M^2W)(x, y)\ dx\, dy = 1$. Note that working with $C(u, v)$ involves a square-root operation: we can work with $C^2(u, v)$ more easily, and it is simple to recover sign information if required. Taking the derivative of $C^2(u, v)$ we find

$$\begin{aligned}\nabla C^2(u, v) = \quad & -2\frac{\iint I(x,y)(MW)(x-u,y-v)dx\, dy}{\iint I^2(x,y)W(x-u,y-v)dx\, dy} \\ \times \quad & \iint I(x, y)\nabla(MW)(x - u, y - v)dx\, dy \\ + \quad & \left[\frac{\iint I(x,y)(MW)(x-u,y-v)dx\, dy}{\iint I^2(x,y)W(x-u,y-v)dx\, dy}\right]^2 \\ \times \quad & \iint I^2(x, y)\nabla W(x - u, y - v)dx\, dy .\end{aligned}$$

This leaves us with just four terms to calculate:

$$\begin{gathered}\iint I(x, y)(MW)(x - u, y - v)dx\, dy \\ \iint I(x, y)\nabla(MW)(x - u, y - v)dx\, dy \\ \iint I^2(x, y)W(x - u, y - v)dx\, dy \\ \iint I^2(x, y)\nabla W(x - u, y - v)dx\, dy .\end{gathered}$$

These equations are easily transformed to the discrete domain, and allow us to compute the correlation gradient at a point with little more than 1/2 the number of operations required to compute the correlation values at 8 neighbours. The values of ∇W and ∇MW can be precomputed when the target is analyzed off-line prior to starting the search. Figure 3 shows these precomputed values for the target shown in Figure 2.

3.3 Performing the Search

The first step in the search is to build pyramid representations for the image to be searched (the pyramid for the target has been done off-line earlier). The depth of this pyramid will be determined by the maximum depth of the target pyramid. The search begins with a full NGC at the between the top level of the image and target pyramids (see Figure 4). The reduced size of the top-level images gives a speed up of $2^{4(k_{\max}-1)}$ over using NGC on the original images (*e.g.* 4096 times faster for a 4-level pyramid). The resulting peaks in the correlation surface are potential matches. Determining the peaks makes use of the worst acceptable score described in section 3.1.1. We assume that prior to starting the search the user specifies an *accept threshold*, and possibly a maximum number of expected targets. While the latter is not necessary for correct operation of the algorithm, it may help speed up the search process even further. Candidates whose final correlation score is below the accept threshold are rejected by the search.

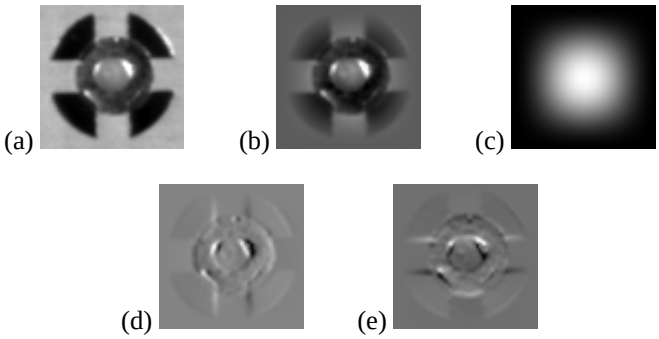

Figure 3. This figure shows the various target representations required at each level of the pyramid for gradient correlation search. The target is given in (a). The windowed version of the target is shown in (b), with the windowing function shown in (c). The horizontal and vertical components of the gradient are shown in (d) and (e).

For each candidate the search descends through the pyramid using a coarse-to-fine search strategy. The candidate's location estimate from the previous level is used as a starting point, and the location estimate in the current level is refined using the correlation gradient descent. When the location estimate for the bottom level of the image has been refined, the search is complete. Any candidate whose bottom level correlation score falls below the threshold is rejected. Candidate targets are also rejected if their contrast is too low as NGC will match anything to a uniform image region.

Finally, the location estimate can be further refined using the subpixel localization algorithm described in Section 3.4. It should be noted that the image can be searched for multiple targets at once. The image pyramid need only be built once, and as many targets as desired can be searched, either sequentially or in parallel.

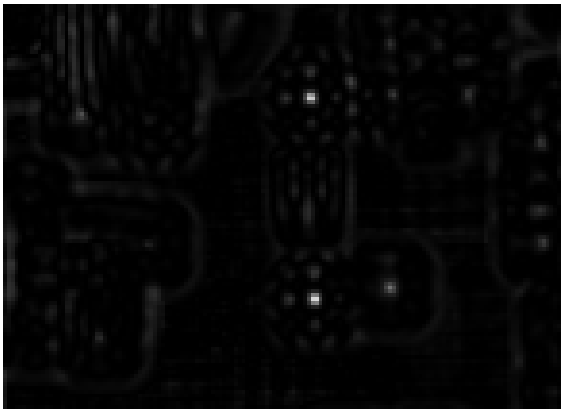

Figure 4. The correlation surface from the top level of the pyramid is shown. Two strong peaks, representing the location of the two instances of the target, are evident. These peaks provide coarse location estimates, which are refined as the algorithm descends through the pyramid.

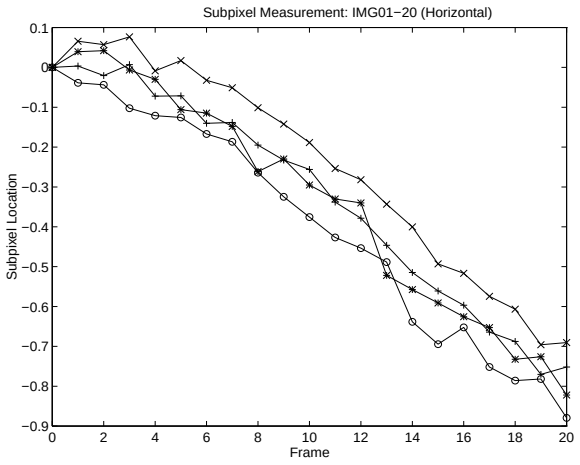

Figure 5. This figure shows the results from doing subpixel estimation in a sequence of images where each image is horizontally offset from the previous one by roughly $1/20$th of a pixel.

3.4 Subpixel Localization

Using a correlation approach to locate a target instance only returns an integral location, *i.e.*, the values of x and y are always integer valued. Assuming that the correlation surface is relatively smooth, it is possible to interpolate between sample values to estimate a sub-pixel localization for the target. We propose a method using a *bi-quadratic surface* for the interpolation.

Once the integral location is found, the 8-neighbours on the correlation surface are also computed. We wish to fit $C(x, y)$ to a function of the form

$$\hat{C}(x, y) = \alpha x^2 + \beta y^2 + \gamma x + \delta y + \xi xy + \varphi$$

where the parameters of the model are

$$\vec{\pi} = [\ \alpha \quad \beta \quad \gamma \quad \delta \quad \xi \quad \varphi\ ]^T\ .$$

Then $\vec{\hat{C}} = F\vec{\pi}$ where F is a co-efficient matrix based solely on the values of x, y, x^2 and y^2 at the sample locations. Since the value of F and C are known, we estimate π using $\vec{\pi} = (F^T F)^{-1} F^T \vec{C}$ where $\vec{C}$ is the value of C at the 9 sample points described by F. All that remains is to find the location of the maximum. At the maximum

$$\begin{aligned} \partial\hat{C}/\partial x = 0 \quad &\Rightarrow \quad 2\alpha x_c + \gamma + \xi y_c = 0 \\ \partial\hat{C}/\partial y = 0 \quad &\Rightarrow \quad 2\beta y_c + \delta + \xi x_c = 0\ . \end{aligned}$$

The subscripts in x_c and y_c indicate those values of x and y which satisfy the requirements for a maximum. This can be rewritten as

$$\begin{bmatrix} 2\alpha & \xi \\ \xi & 2\beta \end{bmatrix} \begin{bmatrix} x_c \\ y_c \end{bmatrix} = - \begin{bmatrix} \gamma \\ \delta \end{bmatrix}$$

with a solution of

$$\begin{bmatrix} x_c \\ y_c \end{bmatrix} = -\frac{1}{4\alpha\beta - \xi^2} \begin{bmatrix} 2\beta & -\xi \\ -\xi & 2\alpha \end{bmatrix} \begin{bmatrix} \gamma \\ \delta \end{bmatrix}\ .$$

It is shown in [9] that RMS error values in the range of 0.1 pixels can be achieved with this technique. Figure 5 shows subpixel localization results from a set of real images where the camera was shifted horizontally by about $1/20$th of a pixel between images using a precision xy translator. Four targets were defined from the first image, and tracked through the sequence.

4 Results

Figure 1 shows a typical search region, while Figure 2 shows a target to be searched for. Figure 4 shows the result of the NGC at the top level of the pyramid, yielding two strong candidates for the search. Both target instances are correctly identified in the image, with a search time on the order of 50 ms using a 200MHz Pentium PC with 64MB of RAM.

Testing with an accept threshold of 0.95, the false-positive rate was small (typically less than 0.1%). In general the false positive rate will vary with the accept threshold, but it should be noted that this allows the search to find very similar objects instead of just identical matches. The false negative rate also

depends on the accept threshold and, if specified, the maximum number of expected targets. This rate rises if the accept threshold is set too high.

It is observed that the time to search for 'large' targets differs little from that for searching for smaller ones, as larger targets tend to result in a deeper pyramid, keeping the top level NGC about the same size. Since usually only 2–3 steps of gradient descent are needed at each level, processing each candidate at each level of the pyramid is fast compared to building the pyramid and doing the top-level NGC.

Search time is lengthened if there are a large number of candidates to check, and thus depends on the relative similarity of background regions to the target. If a maximum number of expected targets is known in advance (as is sometimes the case in industrial applications) this lengthening can be avoided.

5 Discussion

We have described a method for fast pattern search in images that extends NGC to use a pyramid image representation and gradient descent search on the correlation surface. While the results are good, it is important to note that certain limitations exist with this method. First, the technique does not allow for variation in the size or orientation of a target instance with respect to the target image. In practice we have observed that the method is tolerant up to 8–10% of variance, with performance degrading rapidly after that. This is a known limitation of NGC, and is not new to our method. Use of a pyramid representation holds promise for size invariant recognition, since searches can be carried out at different pyramid levels in the search image holding the target pyramid level fixed with little extra cost. Our method differs from [1] in that a simpler pyramid structure is used, and we do not decompose the target into features. Using features might alleviate the occlusion problem, but would detract from speed of operation. Local correlation gradient estimation is also novel compared to [1].

The method also does not handle illumination variation such as cast shadows which were not present when the target was analyzed, or partially occluded target instances. Again, these are limitations inherent in NGC approaches. Finally, instances of pathological image structure which fool NGC, such as searching for line segments, also fool this method. In a case like this, care must be taken to observe the determinant of the matrices used in the subpixel localization lest the matrix become ill-conditioned.

6 Conclusion

In conclusion, we have presented a method for fast pattern recognition and localization in grey-scale images. The technique can be implemented on modest microcomputer systems while still attaining fast response times. The technique is novel in its use of a pyramid image representation to reduce the computational complexity of NGC, and in its use of a local estimate of the gradient of the correlation surface. Also new are the methods used to determine an appropriate depth of the pyramid representation and the subpixel localization algorithm.

7 Acknowledgments

The authors would like to thank Fernando Nuflo and Rob McReady for their assistance with this work.

References

[1] P. Burt. Attention mechanisms for vision in a dynamic world. In *Proceedings of the International Conference on Pattern Recognition*, pages 977–987, 1988.

[2] R. C. Gonzalez and P. Wintz. *Digital Image Processing*. Addison-Wesley Publishing Company, Reading, Massachusetts, 2nd edition, 1987.

[3] A. Goshtasby. Template matching in rotated images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-7(3):338–344, May 1985.

[4] B. K. P. Horn. *Robot vision*. The MIT Press, Cambridge, Massachusetts, 1986.

[5] M. K. Hu. Visual pattern recognition by moment invariants. *IRE Trans. Information Theory*, IT-8:179–187, 1962.

[6] J.-M. Jolion and A. Rosenfeld. *A pyramid framework for early vision*. Kluwer Academic Publishers, P.O. Box 17, 3300 AA Dordrecht, The Netherlands, 1994. ISBN: 0-7923-9402-X.

[7] A. D. Kulkarni. *Artificial Neural Networks for Image Understanding*. Van Nostrand Reinhold, New York, 1994. ISBN 0-442-00921-6; LofC QA76.87.K84 1993.

[8] T. Lindeberg. *Scale-space theory in computer vision*. Kluwer Academic Publishers, P.O. Box 17, 3300 AA Dordrecht, The Netherlands, 1994. ISBN: 0-7923-9418-6.

[9] W. J. MacLean and J. K. Tsotsos. Fast pattern recognition using normalized grey-scale correlation in a pyramid image representation. In progress.

[10] W. K. Pratt. *Digital Image Processing*. John Wiley & Sons, Inc., New York, 2nd edition, 1991.

[11] G. van der Wal and P. Burt. A vlsi pyramid chip for multiresolution image analysis. *Int. Journal of Computer Vision*, 8:177–190, 1992.

[12] H. Wechsler and G. L. Zimmerman. 2-d invariant object recognition using distributed associative memory. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(6):811–821, November 1988.

[13] H. Wechsler and G. L. Zimmerman. Distributed associative memory (dam) for bin-picking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(8):814–822, November 1989.