A novel approach to multi-resolution technique for fast pattern recognition

Xiaochun Liu\textsuperscript{1}, Xiaohua Ding\textsuperscript{2}, Xiang Zhou\textsuperscript{2}, Wen Cui\textsuperscript{3}

\textsuperscript{1} Department of Military Aerospace, College of Aerospace Science and Engineering, National University of Defense Technology, Changsha, 410073, China

\textsuperscript{2} Shenzhen Eagle Eye Online Electronic Technology Co. Ltd, Shenzhen, 518100, China

\textsuperscript{3} 96656 Troops of Rocket Army, Beijing, 100000, China

Abstract. Multi-resolution is a commonly used acceleration algorithm for pattern search. However, there still exist two unsolved questions: “how many levels of image pyramid should be built” and “how to propagate the badly degraded truth patterns to the next pyramid level”. We address these two questions through a novel distinction analysis and propagation strategy based on multimodality analysis. Experimental results show that this method has good practicability and advantage.

1. Introduction

By reducing the area to be searched and the number of pixels used in pattern matching algorithm, multi-resolution technique is able to speed up the computational efficiency of pattern matching dramatically. Multi-resolution technique has been commonly used and also played an important role in many real-time pattern matching applications. Important applications of multi-resolution include image matching [1]-[4], pattern recognition [5]-[8], motion estimation [9]-[11], frame interpolation [12], image registration [13] and Fractal decoding [14]. A multi-resolution technique usually works as follows: Upon building the image pyramids for both the pattern template and searching image, the first search is conducted with the most compressed pattern template and searching image. The resulting location provides a coarse location of the truth pattern template to the next level of the searching image. Therefore, instead of performing a complete search in the next level, one requires to only search a close vicinity of the coarse location obtained from the previous search. This sequence is iterated until the search in the original image is completed. One disadvantage of building image pyramid is the loss of detail information. In some cases, the truth and false corresponding templates only differ in relation to details, which may render the template unrecognizable for template matching algorithms when running at the most compressed level of image pyramid. Thus, a novel distinction analysis is presented to measure the self-distinction between the template and its neighboring templates, and then determine whether to continue building image pyramid or not based on the value of self-distinction. In practical use, the truth corresponding point maybe degrade to the second best matching point (or even worse) because of the loss of detail information, which means the conventional method that propagates only one candidate point may not find the truth point. To address this problem, we present a multimodality detection that can provide a set of rather than only one candidate points to the next level.
The paper is structured as follows. In section 2, the definition and procedures of distinction analysis are described, and some experimental results are also shown to demonstrate our idea. The propagation strategy based on multimodality detection is described, and comparison experiments are also given in section 3. Finally, concluding remarks are provided in section 4.

2. Distinction Analysis

Because of image filtering and scale reduction when building image pyramid, the detail information of pattern template will be inevitably lost, which may render the template too ambiguous to recognize. To answer the question “how many levels should be used to construct the image pyramid to keep the template recognizable”, a novel method that determines the proper number of level according to the characteristics of template as a heuristic manner is described as follows (as shown in Fig.1): (1) conduct self-distinction analysis using the template and its neighboring templates whose centers are equally spaced on a circle of radius R as shown in Fig.2; (2) If the template on the current level is still recognizable, the self-distinction value will be relatively large, and then we can decide to continue building the next image pyramid. Otherwise, end the iteration.

Inspired by the work in the literatures [16, 17], we define the distinction measurement using the template and its neighboring templates. The definition of distinction measurement on the 1st pyramid level goes as follows. (1) As shown in Fig.2, select the center of template as the center of circle, and select other neighboring templates whose centers are equally spaced on the circle of radius R. (2) In this paper, we employ the zero-mean normalized cross-correlation (ZNCC) as the similarity measure function which is a contrast invariant variation. Let f and g denote the template and its neighboring template respectively. The definition of ZNCC can be expressed as follow:

\[
ZNCC(u, v) = \frac{\sum_{x=1}^{m} \sum_{y=1}^{n} (f(x+i, y+j) - \bar{f})(g(u+i, v+j) - \bar{g})}{\sqrt{\sum_{x=1}^{m} \sum_{y=1}^{n} (f(x+i, y+j) - \bar{f})^2} \sqrt{\sum_{x=1}^{m} \sum_{y=1}^{n} (g(u+i, v+j) - \bar{g})^2}},
\]

where: \( \bar{f} \) and \( \bar{g} \) denote the mean value of f and g, \( m \times n \) denotes the size of template, (x, y) and (u, v) are the beginning points of f and g respectively. (3) Using Eq.1, get the ZNCCs between the pattern template and its neighboring templates. The definition of Distinction Measurement (DM) can be expressed as Eq.2,

\[
DM = \frac{1 - \max \left( ZNCC \right)}{2}.
\]
If the pattern template is completely distinctive, the value of DM is equal to 1. If we define $R$ as the radius of the circle on the 1st level, the radius on the $k$th level can be defined as Eq.3,

$$R_k = \frac{R}{2^{k-1}}. \quad (3)$$

From Eq.3, although each level is given a different value of $R$, the equivalent radius keeps invariant for all levels when converting to the 1st level. To save more computational time, the enlarged template instead of the whole image is adopted to calculate DM.

2.1. Experiments of Distinction Analysis

Because Gaussian pyramid is able to efficiently suppress the high-frequency noise caused by pixel interpolation while constructing pyramid and provide much of the information in the coarser scales, in this paper we adopt the Gaussian pyramid rather than four points mean pyramid as multi-resolution image representation [15]. (Our computation is slightly more than four points mean pyramid.)

The results of distinction analysis are shown in Fig.3-4 and Table 1. The pre-set threshold of distinction measurement is given a value of 1/3. How many levels should be used for image pyramid is decided according to the characteristic of the pattern template. For example: the first template including a person is more distinctive than the second template, thus it can afford more levels of image pyramid, and has larger value of distinction (as shown in Tab.1). From Table 1, we can see that the distinction values of the template 1 and 2 all decrease as the increase of pyramid level, which agree with our theoretical analysis above.

3. Propagation Strategy Based On Multimodality Detection

3.1. Theoretical Analysis

The pattern template normally becomes increasingly hard to recognize as the pyramid level increases. In some cases, the resulting location propagated from the most compressed level may be wrong, which will lead the total matching fail because of cascade mode. To address this problem, a multimodality-based propagation strategy is proposed to improve the conventional algorithm to handle multimodality situations that are quite common in practical applications. Actually, the conventional multi-resolution search relies on the “unimodality assumption”. As a result, its applicability is severely limited to the situations when the pattern templates are highly unique. However, it is common that a template has multiple similar candidate templates that only differ in relation to detail information in the most compressed search area, which will produce a correlation matrix with multiple peaks. In this situation,
the truth location may be corresponding to any one of peaks of correlation matrix, which obviously violate “unimodality assumption”. The traditional propagation strategy, which propagates only one candidate location corresponding to the first highest peak, will inevitably cause wrong matching.

As shown in Fig.5, we first choose a template from a visible image, and then conduct the search in an infrared search area. The size of template and search area are 101×101(pixels) and 201×201(pixels) respectively. The correlation matrix is calculated according to ZNCC. The template matching is conducted on different pyramid level successively. From the images of Fig.5, we can see on the 1st level, the truth location corresponds to the first highest peak of correlation matrix since relatively rich detail information. However, as the loss of detail information when constructing pyramid, the truth location begins to degrade (the second highest peak on 2nd level and the third highest peak on 3rd level). To solve the problem, one can propagate those candidate locations whose values of ZNCC are larger than a pre-set value, and then conduct the search at the vicinity of those locations on the finer pyramid level [5]. This approach may have better performance as more candidate locations have been provided. However, there exist two main problems in this solution. If the first highest peak that may correspond to false location is relatively strong, the candidate locations will gather at its vicinity, which may lead the truth location fail to be propagated. In addition, the approach often provides too many candidate locations, and increases the computational time too much.

Fig. 3. Distinction analysis for the pattern template1 and pattern template 2

In this work, a new propagation strategy based on “multimodality assumption” is presented to find the truth location that may degrade. Our method begins with a reasonable assumption that the truth template is still one of templates similar to the template no matter how it has degraded. Because there normally exists a peak that is formed by the similar template in the correlation matrix, in order to propagate truth locations, we just need find all locations of peaks. Hence, the multimodality detection is presented to detect peaks and find their locations in the correlation matrix. The value of peak is normally higher than its neighbors, which means it corresponds to a local maximum value. To find those peaks, each value of correlation matrix is compared to its neighbors within a circle of radius R in the correlation matrix. It is selected only if it is larger than all of these neighbors. If we define the ZNCC(i, j) and ZNCC(i+m, j+n) are the value of location(i, j) and (i+m, j+n), the multimodality detection can be expressed as follow:

\[ ZNCC(i, j) \geq ZNCC(i+m, j+n), \quad (\sqrt{m^2 + n^2} \leq R). \]  (4)
Fig. 4. Examples for constructing image pyramid based on distinction analysis. The images in the left and right rectangle are the results for the first and second pattern template respectively. For each group, upper row is the Gaussian pyramid for each level. Bottom row is the Gaussian pyramid expanded to the size of the original image.

To save computational time, one can choose a pre-set threshold (CT) to constrain the number of candidate locations. Let \( C_{\text{max}} \) denote the maximum value of correlation matrix, \( PVR \) denote the Peak Value Ratio, then a threshold namely \( CT \) can be defined as follow:

\[
CT = C_{\text{max}} \times PVR.
\] (5)

One doesn’t need to reset \( CT \) when the template and/or search area changes, because the value of the \( CT \) can be adjusted according to the characteristic of correlation matrix. One can also set a threshold (NT) to constrain the number of candidate locations. We first need to sort the value of ZNCC at the candidate locations as a descending order, and then a location is selected only if its value of ZNCC ranks among the top NT. Using this Method, we can make sure the candidate locations and the peaks are one-to-one correspondences, thus we can avoid those candidate locations gathering at vicinity of some strong peak. In this work, the radius \( R \), the Peak Value Ratio \( PVR \) and the threshold of the number are experimentally given a value of 5, 0.75, and 3 respectively.

3.2. Experiments for Multimodality Detection
We have evaluated the performance of the proposed algorithm using three multi-sensor image pairs and many pattern matching experiments. Because in our experiments all the truth locations are not corresponding to the highest peak of correlation matrix on the most compressed level, the traditional method, which propagates only the highest peak, cannot handle this situation. Therefore, we only need to compare the results obtained from the proposed algorithm with those from previous method presented in paper [5]. In order to justify the improved propagation strategy based on the multimodality detection, the two methods were both used to get the number and locations of similar templates for each experiment with the predetermined \( PVR \) or NT respectively. The experimental results are shown in the table 2 and 3. The matching images were well registered before the template matching. The original size of template is 101×101(pixels).
Fig. 5. The results of template matching on different levels of Gaussian pyramid. The top row is the visible pattern image (left) and the infrared search image (right), and the resulting locations for different levels are denoted in the search image. The results of different pyramid levels are shown from the second row to the fourth row. In each row, from left to right: (1) pattern template which are expanded to the original resolution, (2) the three dimensional correlation matrix, (3) the correlation matrix viewed from X to Y, (4) the resulting template.

From table 2, we can get the conclusion that: for the given pre-set threshold NT, the candidate locations obtained from the previous method gather at the vicinity of some false but strong peaks; the truth location is almost not propagated; however, because the candidate locations obtained from the proposed method are a one-to-one correspondence with the peaks of correlation matrix, it can propagate the truth location robustly. From table 3, we can get the conclusion that: for the given pre-set threshold PVR, both methods are able to propagate the truth location, but the number of candidate locations obtained from our method is much less than the previous method. The mean number of our method is 3.4, whereas the number of the previous method is 155.7. Apparently, our method can save more computational time. Note that: only one failure occurred because the truth location has degraded to the sixth highest peak. If NT is given a value of 6, we will also propagate the truth location correctly.

Table 2. Comparisons between our method and previous method when NT = 3.

| Image pair | Candidate number | True/False | Candidate number | True/False |
|------------|------------------|------------|------------------|------------|
| 1          | 3                | True       | 3                | False      |
| 1          | 3                | True       | 3                | False      |
| 2          | 3                | False      | 3                | False      |
| 2          | 3                | True       | 3                | False      |
| 2          | 3                | True       | 3                | False      |
| 3          | 3                | True       | 3                | True       |
| 3          | 3                | True       | 3                | False      |
4. Conclusion

A novel approach to multi-resolution technique for fast template matching is proposed in this paper. There are two main contributions in this work. (1) We present a distinction analysis to decide how many levels should be built in the image pyramid according to the characteristics of the pattern template. (2) We present a propagation strategy based on multimodality detection. Using this strategy, we can propagate the truth location even if it has degraded badly. In addition, through two thresholding schemes, we can constrain the number of candidate locations. The experimental results demonstrate that our method can propagate the truth location efficiently and robustly even in the challenging situation, and has a significant advantage over the previous methods.

References

[1] Sabyasachi Pal, Titas De, Mainak Sen, Arnab Chatterjee: Grey-Scale Template Image Matching by Area Based Matching Techniques. In IECON 2011 organized in collaboration with IEEE on 5th&6th, (2011).

[2] E H Adelson, C H Anderson, J R Bergen, P J Burt, and J M Ogden: Pyramid methods in image Processing. RCA Engineer, vol. 41, pp. 29-33, (1984).

[3] Ryo Takei: A New Grey-Scale Template Image Matching Algorithm Using the Cross-Sectional Histogram Correlation Method. (2003).

[4] Lin, S.J., Chung, K.L., Chang, L.C.: An improved search algorithm for vector quantization using mean pyramid structure. Pattern Recognition Letter. 22 (3/4), pp.373–379(2001).

[5] Jingge Wu, Jordan Kuhn: Pattern matching method using a pyramid structure search, image detecting device, and computer-readable medium. United States Patent, Patent No.: 7215816B2 (2007).

[6] Sheikholeslami G., Chatterjee S., Zhang A. Wave Cluster: A Multi-Resolution Clustering Approach for Very Large Spatial Databases. In Proc. 24th Int. Conf. on Very Large Data Bases, New York, NY, pp. 428 – 439, (1998).

[7] W.J. MacLean and J.K.T. Tsotsos: Fast pattern recognition using gradient-descent search in an image pyramid. In Proceedings of the 15thInternational Conference on Pattern Recognition, pp. 877-881(2000).

[8] Lee, C.H., Chen, L.H.: A fast search algorithm for vector quantization using mean pyramids of code word. IEEE Transactions on Consumer Electronics, vol. 43 (2/3/4), pp.1697–1702(1995).

[9] Jong-Nam Kim and Tae-Sun Choi: Adaptive matching scan algorithm based on gradient magnitude for fast full search in motion estimation. IEEE Transactions on Consumer Electronics, vol. 45(3), pp. 762 – 772(1999).

[10] G. de Haan: Progress in motion estimation for video format conversion,” IEEE Transactions on Consumer Electronics, vol. 46, pp. 449–459(2000).

[11] B Jahne, H Haussecker, and P Geissler: Bayesian Multi-Scale Differential Optical Flow. Handbook of Computer Vision and Applications, vol. 2, chapter 14, pp. 297-422, Academic Press (1999).

| Image ID | Candidate number | True/false (TMRT) | True/false | True/false |
|----------|------------------|-------------------|-------------|-------------|
| 1        | 1                | True              | 658         | True        |
| 1        | 4                | True              | 92          | True        |
| 2        | 6                | True              | 71          | True        |
| 2        | 2                | True              | 14          | True        |
| 2        | 4                | True              | 158         | True        |
| 3        | 2                | True              | 57          | True        |
| 3        | 2                | True              | 40          | True        |
[12] B.W. Jeon, G.I. Lee, S.H. Lee and R.H. Park: Coarse-To-fine Frame Interpolation for Frame Rate Up-Conversion Using Pyramid Structure. IEEE Transactions on Consumer Electronics, vol. 49(3), pp. 499-508 (1997).

[13] Philippe Thévenaz, Urs E. Ruttimann, and Michael Unser: A Pyramid Approach to Subpixel Registration Based on Intensity. IEEE Transactions on Image Processing, vol. 7(1), pp. 27-41(1998).

[14] Hyun-Soo Kang, Seong-Dae Kim: Fractal decoding algorithm for fast convergence. Optical Engineering, vol. 35(11), pp.3191-3198(1996).

[15] Burt and Adelson: The Laplacian Pyramid as a Compact Image Code. IEEE Transactions on Communications, vol. COM-31, no. 4, pp. 532-540(1983).

[16] John E. Hutchinson: Fractals and Self Similarity. Indiana University Mathematics Journal, Vol.30 (5), pp. 713-747(1981).

[17] Mark E. Crovella and Azer Bestavros: Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes. IEEE/ACM Transactions on Networking, vol. 5(6), pp. 835-846 (1997).