THREE BIRDS, ONE STONE: SIMULTANEOUS OBJECT DETECTION, RECOGNITION, AND PROFILING USING PHASE ENCODED MACE FILTERS

Chandrasekhar Bhagavatula, Felix Juefei-Xu, Jason Wang, and Marios Savvides

Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA

ABSTRACT

In this work, we have proposed a novel design extension for the Minimum Average Correlation Energy (MACE) filter. By encoding transformation information of the object into the phase angle of a complex filter, the proposed Phase MACE (PMACE) filter is able to handle simultaneous object detection, recognition, and profiling, whereas the traditional MACE filter is only able to handle the first two simultaneously. The PMACE can not only accurately detect and recognize the object, but also accurately profile which transformation the object is undertaking, like rotation angle or scaling factor, etc. We have shown the validity of the proposed method through the MNIST handwritten digits database. The PMACE filter constantly shows better detection and recognition performance than MACE filter, while being able to accurately perform profiling.

Index Terms— MACE Filter, Correlation Filter

1. INTRODUCTION

Many tasks in computer vision require detection of a desired object as well as recognition of what that object is. While there have been great advances recently in this area, many of the new techniques require large amounts of training data and lots of computing power in order to train the detectors. However, there are still many scenarios where computing power is restricted and learning may even need to be done on the fly. For example, when using home security cameras, on board processing of the video feed can reduce the amount of data that needs to be stored, reducing storage costs or the need for potentially insecure network connections.

One area of study that has tackled the problem of simultaneous detection and recognition for many years is that of advanced correlation filters. These filters were designed to perform detection of specified objects using a simple correlation. Since correlations can be performed extremely efficiently in the frequency domain, these approaches were very fast. However, most of these approaches suffer from the fact that they cannot distinguish between different variations of the same object such as rotation or scale. While they can be trained in such a way to recognize the object under the different variations, it is possible that the end user would like to know which variation the filter has found. We propose a novel extension that can be applied to advanced correlation filter techniques that allow for an encoding of this information into a complex correlation surface. We show our approach specifically with the Minimum Average Correlation Energy (MACE) filter [1].

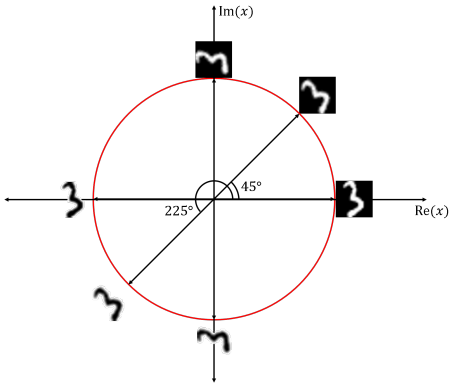

Fig. 1: Example embedding of rotated samples onto the unit circle in complex space. The negative versions of the samples are implicitly embedded on the opposite side.

Related Work: Advanced correlation filter theory mainly begins with the introduction of the Synthetic Discriminant Function (SDF) filter [2]. This is one of the first composite correlation filters, designed to handle general types of distortions by training on several samples at once. The SDF filter is a very simple filter that guaranteed the correlation output at the origin is a specified value. However, this leads to correlation surfaces with very broad peaks and sometimes even values higher at offset locations. The MACE filter [1] is the next filter designed to address this problem. MACE filters maintain the constraint on the origin of the correlation surface but minimize the energy of the surface as well. This tends to put all other values to 0 and create very sharp peaks. Many other types of correlation filters have been developed over the years from filters designed to give optimal performance in the presence of noise [3], filters that trade-off between multiple criteria [4], to even unconstrained filters [5]. Even the problem of circular vs linear correlation in the frequency domain has been addressed recently to give better performing filters [6]. Correlation filters have been used in many detection and recognition tasks. Very accurate eye detection can be accomplished through the use of Average Synthetic Exact Filters

(ASEF) [7]. Open set face recognition using banks of advanced correlation filters has achieved very high performing verification rates [8]. Action recognition in video using 3D correlation filters on temporal data can be done [9]. Even the combination of Support Vector Machine (SVM) and correlation filter theory has led to more generalized filters [10, 11]. However, in all of these approaches, the desired output and the resulting filters are real valued when there is no constraint demanding such a solution. It is from that observation that we propose to use phase encoding filters to distinguish multiple instances of the same class in one correlation filter.

2. PROPOSED METHOD

Traditionally, correlation filters have been designed in the frequency domain to take advantage of the fact that correlation in the time or space domain is conjugation and element-wise multiplication in the frequency domain. This allows for a very simple closed form solution in the case of many filters such as MACE filters. MACE filters aim to minimize the Average Correlation Energy (ACE) of the resulting correlation surface while maintaining a constrained value at the origin of the surface. The combination of these two results in sharp peaks that make detection easy. For any group of samples in the frequency domain, the ACE with a given correlation filter $\mathbf{h}$ is given as $\text{ACE} = \mathbf{h}^+ \left[\frac{1}{d \cdot N} \sum_{i=1}^{N} \mathbf{D}_i \mathbf{D}_i^*\right] \mathbf{h} = \mathbf{h}^+ \mathbf{D} \mathbf{h}$, where d is the dimension of the samples, N is the number of samples, and $\mathbf{D}_i$ is a diagonal matrix with the i^{th} sample along the diagonal. The origin constraints can be easily expressed in the frequency domain as $\mathbf{X}^+\mathbf{h} = d \cdot \mathbf{u}$, with $\mathbf{X}$ being a matrix with the vectorized data samples along the columns and $\mathbf{u}$ being a vector containing the desired values at the origin, usually 1 for positive samples and 0 for negative. By minimizing the ACE term subject to the set of constraints, the solution for $\mathbf{h}$ can be found as $\mathbf{h} = \mathbf{D}^{-1}\mathbf{X}\left(\mathbf{X}^+\mathbf{D}^{-1}\mathbf{X}\right)^{-1}\mathbf{u}$, and while $\mathbf{D}$ is a large matrix , $d \times d$, it is also diagonal meaning that the computation of $\mathbf{D}^{-1}$ can be done very efficiently. Even $\mathbf{X}^+\mathbf{D}^{-1}\mathbf{X}$ is only an $N \times N$ matrix which is usually small in correlation filter applications.

Though this is how MACE filters are traditionally trained [12, 13, 14, 15], there is no constraint on the formulation of the filter that specifies the constraints must be real. If the constraints are real, the resulting filter will also be real. However, we can specify complex valued constraints which will result in a complex filter. Adding this second dimension to the output allows us to incorporate new behavior into the correlation filters. Specifically, we can now encode more information into the phase of the complex value and leave the magnitude as one. Since correlation filters are usually used for simultaneous detection and recognition, we still want the magnitude of the resulting surface to contain sharp peaks at the locations of objects of interest. However, the phase information at these locations can tell us specifically which objects we are looking at.

We can only encode samples onto a 180° arc of the unit circle however. This is due to the fact that this is a linear method meaning that when we encode one sample at a specific angle on the unit circle, we implicitly encode the negative of that sample on the other side of the unit circle as shown in Figure 1. This is due to the fact the negating a sample results in a negation of the entire correlation output and that for a complex value x, we have $\angle(-x) = 180^\circ + \angle(x)$.

Different classes can be encoded onto the unit circle but the filter will learn some representation that relates samples embedded at closer angles as being more related to each other. Therefore, using this kind of method makes more sense when the samples being embedded have a relation that is easily defined. Two possible variations that could be learned are rotations and scale. By applying these transformations to training samples, a Phase MACE (PMACE) filter could be created to detect these objects in the presence of these variations as well as determine by how much an object was rotated or scaled. Traditional correlation filters would have to learn one filter for each variation in order to determine what rotation or scale was being seen and there would be no relation between them meaning unseen variations would not be classified well. This phase encoding method can easily be extended to other constrained advanced correlation filters such as Optimal Tradeoff Synthetic Discriminant Function (OTSDF) filters [4].

3. EXPERIMENTS

In this section, we will showcase how a single PMACE filter can simultaneously detect, recognize the object and determine the rotation or scaling of the object.

Database: The MNIST database of handwritten digits [16] contains a training set of 60,000 and a testing set of 10,000 32×32 gray-scale images showing hand-written digits from 0 to 9. The original black and white images from NIST were size normalized to fit in a 20×20 pixel box while preserving their aspect ratio. The MNIST dataset is commonly used as a digits classification dataset, while in this work, we use it for simultaneous detection, recognition, and profiling purposes. In the following experiments, we use the normalized 20×20 images.

Experimental Setup: The traditional MACE filter already achieves simultaneous detection and recognition. In this work, we are pushing it one step further, equipping MACE filter with additional profiling capability, determining what transformation the object is undertaking, in addition to the detection and the recognition of the object.

As discussed in the previous section, we need to encode the parameterized transformation information into the phase angle of the PMACE filter. That being said, this work only deals with transformations that can be parameterized. We can thus map the parameters $\mathbf{w}$ of the transformation to the phase angle θ, through some function f. It is important to

Digits												Avg
Rotation GT (°)	0	3	6	9	12	15	18	21	24	27	30	-
Rotation Detected (°)	6.8e-15	3.0079	5.9186	8.6733	12.3456	15	17.7058	20.7493	23.6562	27.4783	30	-
Absolute Error (°)	6.8e-15	0.0079	0.0814	0.3267	0.3456	0	0.2942	0.2507	0.3438	0.4783	0	0.1935
PSR for MACE	194.26	252.35	259.13	251.46	231.39	219.15	230.17	216.90	213.45	220.51	179.28	224.37
PSR for PMACE	210.41	268.51	272.71	262.51	242.18	231.64	241.61	228.00	226.51	235.85	189.87	**237.25**

Table 1: Numerical results of detection, recognition, and profiling for rotation on one object which undertakes rotation transformations.

Digits											
# of Images	1135	1032	1010	982	892	958	1028	974	1009	980	Global Avg
Avg Abs Error	0.1685	0.1333	0.1251	0.1308	0.1262	0.1208	0.1253	0.1230	0.1101	0.1589	0.1322
Std	0.2219	0.1522	0.1453	0.1599	0.1471	0.1454	0.1549	0.1445	0.1351	0.1786	-
Avg PSR for MACE	80.01	182.36	285.01	287.04	347.42	249.86	223.17	227.09	387.15	234.20	250.3310
Avg PSR for PMACE	83.15	198.80	314.51	314.20	384.38	274.49	239.03	249.97	429.22	261.01	**274.8760**

Table 2: Numerical results of detection, recognition, and profiling for rotation on all the 10,000 objects from MNIST which undertake rotation transformations.

note that once we have encoded such information during the filter training stage, we need to decode the phase angle and get the transformation parameters back for the testing cases, in order to profile which transformation the object is undertaking. Therefore, the mapping function f has to be 1-to-1. In our experiments, we will be dealing with transformations such as rotation and scaling, whose transformation parameters $\mathbf{w}$ are thus one dimensional: rotation angle, and scaling factor. For the sake of simplicity, we choose the mapping function f to be linear.

Next, we will briefly go over the metrics to be used for the experiments. Since only the training data is constrained to have the origin at a specific magnitude, a simple threshold often will not produce good results when using correlation filters. Instead the Peak-to-Sidelobe Ratio (PSR) is more often used as a measure of peak sharpness and used to determine where the filter detects an object. The PSR at a location x is computed as $\mathrm{PSR}(x) = \frac{c(x)-\mu_{s(x)}}{\sigma_{s(x)}}$, where $c(x)$ is the value of the correlation surface at x and $s(x)$ is the sidelobe region around x. Generally the sidelobe region excludes some small area immediately surrounding x. In our experiments, the sidelobe region consists of the values greater than 4 and less than or equal to 8 pixels away from x.

After the detection is carried out, we can measure the quality of the learned filter by looking at the absolute error between the ground truth profile and the detected profile.

We will carry out two sets of experiments targeting simultaneous object detection, recognition, and profiling for (1) rotation and (2) scaling. During the filter training stage, each object (digit image) will undertake a set of transformations $\{\mathcal{T}_1\}$, which is also the set for the associated parameters. The PMACE filter will encode the transformation information of this object into its phase angle. During testing, the same object undertakes another set of transformations $\{\mathcal{T}_2\}$, and the trained PMACE filter is expected to decode the transformation information based on the phase angle obtained from the filter response. It is worth noting that the set $\{\mathcal{T}_1\}$ and $\{\mathcal{T}_2\}$ may or nor may not share common elements. What is important is that $\{\mathcal{T}_1\}$ and $\{\mathcal{T}_2\}$ share the same orbit. Specifically, for the rotation experiments, the parameter set for the PMACE training is $\{\mathcal{T}_1\} = \{0°, 5°, 10°, 15°, 20°, 25°, 30°\}$ and the set for the testing is $\{\mathcal{T}_2\} = \{0°, 3°, 6°, 9°, 12°, 15°, 18°, 21°, 24°, 27°, 30°\}$. For the scaling experiments, the parameter set for the PMACE training is $\{\mathcal{T}_1\} = \{0.4, 0.6, 0.8, 1.0\}$ and the set for the testing is $\{\mathcal{T}_2\} = \{0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$. The baseline method will be based on traditional MACE filter that is trained on the same set of object transformations as the PMACE filter. Since MACE filter can only simultaneously detect and recognize, but not profile the object, we will only look at the PSR value of the MACE filter as our benchmark.

Experimental Results on Simultaneous Object Detection, Recognition, and Profiling for Rotation: In the first set of experiment, we will perform simultaneous object detection, recognition, and profiling for rotation using the proposed PMACE filter. We randomly select an object from the MNIST database, and create the rotated versions of the object for both the training and testing scenarios according to the rotation parameters listed in the set $\{\mathcal{T}_1\}$ and $\{\mathcal{T}_2\}$ as discussed above. We then paste the testing objects with transformations $\{\mathcal{T}_2\}$ onto a 400×400 black blanket at random locations. The blanket is convolved with both the PMACE and MACE filters which are trained on the same object with transformations $\{\mathcal{T}_1\}$. The PSR is computed over the entire response map, and we show detection bounding boxes where the PSR values are above some pre-defined threshold determined by cross-validation in the preliminary experiments. The locations of these PSR peaks are where the filters think the transformed objects are located. The phase angles can be easily obtained from the phase surface (based on the real and imaginary response maps), and the actual rotation profiling of the detected object can be decoded accordingly.

Figure 2(L) and (M) show the detected bounding boxes using both MACE and PMACE filters with the PSR values. Figure 2(R) shows the profiling for rotation in the green texts and the ground truth rotation angles in red texts. The numerical results of this experiment are consolidated in Table 1. As can be seen, the average absolute error is $0.1935°$, which is a pretty low profiling error, especially given that only $0°, 15°, 30°$ are seen during the training stage. The last two rows of Table 1 show the PSR values for both the MACE

Digits								Avg
Scale GT (°)	0.4	0.5	0.6	0.7	0.8	0.9	1.0	-
Scale Detected (°)	0.4	0.44477	0.6	0.69765	0.8	-	1.0	-
Absolute Error (°)	0	0.0552	0	0.0023	0	-	0	0.0096
PSR for MACE	859.20	351.39	445.49	326.54	310.71	-	241.77	422.52
PSR for PMACE	895.72	366.01	460.14	337.60	324.84	-	250.31	**439.12**

Table 3: Numerical results of detection, recognition, and profiling for scaling on one object which undertakes scaling transformations.

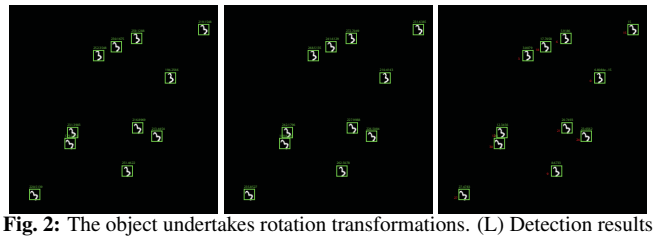

Fig. 2: The object undertakes rotation transformations. (L) Detection results using MACE filter with PSR values; (M) detection results using PMACE filter with PSR values; (R) detection results using PMACE filter with profiled rotation angle (in green) and ground truth angle (in red).

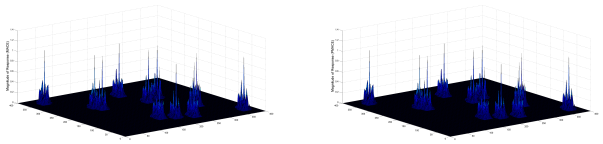

Fig. 3: Response magnitude surface for MACE (L) and PMACE (R) filters.

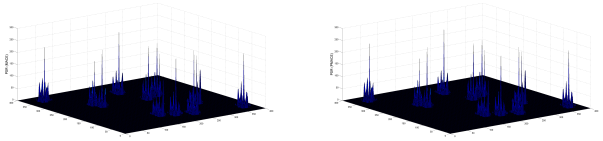

Fig. 4: PSR surface for MACE (L) and PMACE (R) filters.

and PMACE. Under the same training and testing settings, the proposed PMACE filter is able to detect and recognize the object better than the traditional MACE filter by obtaining higher PSR values. Figure 3 shows the response magnitude surfaces for both MACE (L) and PMACE (R) filters. Figure 4 shows the PSR surfaces for both MACE (L) and PMACE (R) filters. Visually, both MACE and PMACE are able to produce sharp peaks at the correct locations, indicating success in both the detection and the recognition of the objects.

Next, we repeat the same experiments for the entire MNIST testing set which contains 10,000 digits. The results are tabulated in Table 2. The first row shows the average image of the digits. The third and fourth row show the average absolute error between the profiled rotation angle and the ground truth, and their standard deviation. The final two rows show the PSR values using both MACE and PMACE filters. As can be seen, the PMACE filter is able to profile the rotation angle at very high accuracy, with a global average error of 0.1322°. Also, the PMACE shows higher PSR values than MACE across all 10 sets of digits.

Experimental Results on Simultaneous Object Detection, Recognition, and Profiling for Scaling: In this section, we extend the idea from rotation to scaling transformations. The objects now undertake scaling with the aforementioned transformation parameter sets $\{\mathcal{T}_1\}$ and $\{\mathcal{T}_2\}$. We again pick

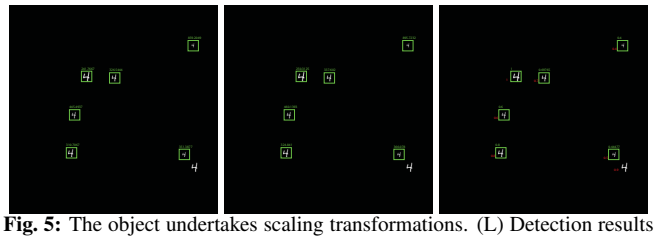

Fig. 5: The object undertakes scaling transformations. (L) Detection results using MACE filter with PSR values; (M) detection results using PMACE filter with PSR values; (R) detection results using PMACE filter with profiled scaling factor (in green) and ground truth scaling factor (in red).

a random object from the MNIST dataset. Figure 5(L) and (M) show the detected bounding boxes using both MACE and PMACE filters with the PSR values. Figure 5(R) shows the profiling for scaling in the green texts and the ground truth scaling factors in red texts. In this particular case, both filters miss one bounding box, so there is one false reject.

The numerical results of this experiment is consolidated in Table 3. As can be seen, the average absolute error is 0.0096, which is again a pretty low profiling error. The last two rows of Table 3 show the PSR values for both the MACE and PMACE. Under the same training and testing settings, the proposed PMACE filter is able to detect and recognize the object better than the traditional MACE filter by obtaining higher PSR values. We repeat the same experiments for the entire MNIST testing set and obtain similar results. Due to space constraints, we have to omit the table.

4. CONCLUSIONS

In this work, we have proposed a novel design extension for the MACE filter. By encoding transformation information of the object into the phase angle of a complex filter, the PMACE is able to handle simultaneous object detection, recognition, and profiling, whereas the traditional MACE filter is only able to handle the first two simultaneously. The PMACE can not only accurately detect and recognize the object, but also accurately profile which transformation the object is undertaking, like rotation angle or scaling factor, etc. We have shown the validity of the proposed method through the MNIST handwritten digits database. The PMACE filter constantly shows better detection and recognition performance, via PSR values, than MACE filter, while being able to accurately perform profiling. In the future, we would like to explore how this method performs with other constrained correlation filters as well as how it can be extended to unconstrained filters.

5. REFERENCES

[1] Abhijit Mahalanobis, B. V. K. Vijaya Kumar, and David Casasent, "Minimum average correlation energy filters," *Appl. Opt.*, vol. 26, no. 17, pp. 3633–3640, Sep 1987.

[2] Charles F. Hester and David Casasent, "Multivariant technique for multiclass pattern recognition," *Appl. Opt.*, vol. 19, no. 11, pp. 1758–1761, Jun 1980.

[3] B. V. K. Vijaya Kumar, "Minimum-variance synthetic discriminant functions," *J. Opt. Soc. Am. A*, vol. 3, no. 10, pp. 1579–1584, Oct 1986.

[4] J. Figue and Ph. Réfrégier, "Optimality of trade-off filters," *Appl. Opt.*, vol. 32, no. 11, pp. 1933–1935, Apr 1993.

[5] Abhijit Mahalanobis, B. V. K. Vijaya Kumar, Sewoong Song, S. R. F. Sims, and J. F. Epperson, "Unconstrained correlation filters," *Appl. Opt.*, vol. 33, no. 17, pp. 3751–3759, Jun 1994.

[6] J. A. Fernandez and B. V. K. Vijaya Kumar, "Zero-aliasing correlation filters," in *2013 8th International Symposium on Image and Signal Processing and Analysis (ISPA)*, Sept 2013, pp. 101–106.

[7] D. S. Bolme, B. A. Draper, and J. R. Beveridge, "Average of synthetic exact filters," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, June 2009, pp. 2105–2112.

[8] Chunyan Xie, M. Savvides, and B. V. K. Vijaya Kumar, "Redundant class-dependence feature analysis based on correlation filters using frgc2.0 data," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops*, June 2005, pp. 153–153.

[9] M. D. Rodriguez, J. Ahmed, and M. Shah, "Action mach a spatio-temporal maximum average correlation height filter for action recognition," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, June 2008, pp. 1–8.

[10] A. Rodriguez, V. N. Boddeti, B. V. K. Vijaya Kumar, and A. Mahalanobis, "Maximum margin correlation filter: A new approach for localization and classification," *IEEE Transactions on Image Processing*, vol. 22, no. 2, pp. 631–643, Feb 2013.

[11] Vishnu Naresh Boddeti and B. V. K. Vijaya Kumar, "Maximum margin vector correlation filter," *CoRR*, vol. abs/1404.6031, 2014.

[12] F. Juefei-Xu, K. Luu, and M. Savvides, "Spartans: Single-sample Periocular-based Alignment-robust Recognition Technique Applied to Non-frontal Scenarios," *IEEE Trans. on Image Processing*, vol. 24, no. 12, pp. 4780–4795, Dec 2015.

[13] F. Juefei-Xu and M. Savvides, "Subspace Based Discrete Transform Encoded Local Binary Patterns Representations for Robust Periocular Matching on NIST's Face Recognition Grand Challenge," *IEEE Trans. on Image Processing*, vol. 23, no. 8, pp. 3490–3505, aug 2014.

[14] D. K. Pal, F. Juefei-Xu, and M. Savvides, "Discriminative Invariant Kernel Features: A Bells-and-Whistles-Free Approach to Unsupervised Face Recognition and Pose Estimation," in *Computer Vision and Pattern Recognition (CVPR), 2016 IEEE Conference on*, June 2016.

[15] P. Buchana, I. Cazan, M. Diaz-Granados, F. Juefei-Xu, and M.Savvides, "Simultaneous Forgery Identification and Localization in Paintings Using Advanced Correlation Filters," in *IEEE International Conference on Image Processing (ICIP)*, Sept 2016, pp. 1–5.

[16] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.