SOFT SEGMENTATION WEIGHTED IECO DESCRIPTORS FOR OBJECT RECOGNITION IN SATELLITE IMAGERY

Stanton R. Price, Derek T. Anderson

Mississippi State University
Electrical and Computer Engineering Department
Mississippi State, Mississippi, USA

Matthew R. England, Grant J. Scott

University of Missouri
Center for Geospatial Intelligence
Columbia, Missouri, USA

ABSTRACT

Object recognition from remote sensing systems is a task of immense interest. With the vast deployment of aerial vehicles and space borne sensors for a wide variety of purposes, it is critical to have robust image processing techniques to analyze massive streams of collected data. Herein, we explore the utility of a feature descriptor learning framework, called *improved Evolution-COnstructed* (iECO) features. Additionally, an investigation into the combination of iECO features with soft features is conducted. Soft features are a deterministic approach to highlighting pertinent information for improving the quality of features extracted specific to the object of interest while iECO is a way to learn from data the relevant information. Experiments are conducted using four-fold (scene based) cross-validation and are reported in terms of target recognition rates and false alarm rates. Results indicate that iECO features are individually best overall and the combination of iECO and soft features can lead to improved results.

***Index Terms*—** iECO, feature learning, soft features, object recognition

1. INTRODUCTION

As the popularity and deployment of remote sensing systems continues to rise, so does the demand for signal and image processing of single-, multi-, and hyper-spectral big data. Applications of remote sensing range from military and homeland defense, e.g., person, car, tank or airplane detection and tracking, to earth observations like the identification and monitoring of crops, vegetation, and deforestation. While applications vary, most demand the robust detection or recognition of anthropogenic or natural phenomena. Rather than develop overly specific algorithms for each application or object, we instead seek new flexible and generalized processing frameworks that can be tailored to different needs. In this article, we explore the individual and combined benefits of a recently established feature descriptor learning framework called *improved Evolution-COnstructed* iECO [1], initially explored in the context of explosive hazard detection for humanitarian mine clearance, and a multi-scale importance map weighted approach for soft feature extraction [2]. The prior approach is ideal for learning signal features while the latter is of benefit for soft object detection and the extraction of features in rich foreground (target) regions. The objective is to combine these two methods in order to learn signal features in target/object rich *regions of interest* (ROI). Figure 1 is a flow diagram highlighting the proposed methods (expanded on in Section 2).

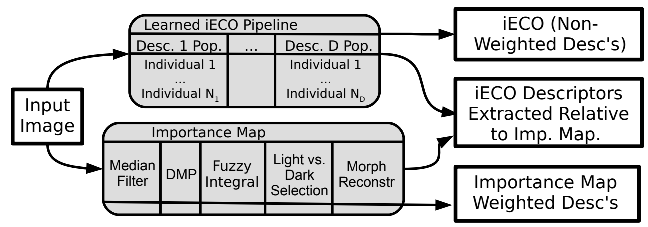

Fig. 1. Flow diagram of the proposed work. The goal is to investigate the individual and combined performance of iECO and soft features in remotely sensed data.

In order to demonstrate the benefit and performance of the proposed methodology, we study the application of automatic aerial vehicle recognition in satellite imagery. Our image database consists of panchromatic high-resolution orthorectified, georeferenced commercial satellite imagery from DigitalGlobe's *Quickbird* sensor. This database consists of numerous objects from four scenes representing different capture times of a single region within the four scenes, Kabul International Airport. Figure 2 shows example target imagery. The remainder of this article is organized as follows. In Section 2 we discuss methods and new work, and in Section 3 we discuss findings.

2. METHODS

Before we explain iECO, importance maps, and how they are combined, the general architecture of our approach is outlined. First, this work is one of sliding window-based object recognition. Feature extraction, relative to a single chip,

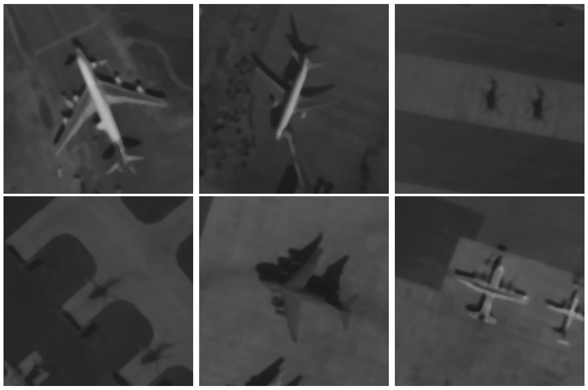

Fig. 2. Example objects from our image database, which shows the variations within and across object classes and complicated context. From top left, moving right: two commercial jets instances, two helicopter instances, and two military/cargo instances.

called ROI hereafter, is performed across multiple spatial resolution scales under a pyramid scheme. In this work, we used 0.5$\times$0.5 m *ground sample distance* (GSD) images, as well as downsampled resolutions of 1×1 m and 2×2 m GSD. Furthermore, a 5×5 partially overlapping cell-structured approach is implemented at each scale. This processing technique is important as it helps preserve the local spatial relations of features in an image ROI. Features from cells and scales are ultimately concatenated into a unified vector, and *support vector machine* (SVM) classification is used. Next, we discuss the iECO framework.

As detailed in [1], the iECO feature descriptor framework, referred to hereafter as simply iECO, improves the ECO work of Lillywhite et al. [3] in two major overarching ways; replacing unrolled image "features" with feature descriptors and more efficient and effective learning. The goal of iECO is to learn a composition of image transforms using some learner, herein a *genetic algorithm* (GA), such that some defined image space feature descriptor can best extract discriminative information. That is, iECO provides an autonomous mechanism to potentially enhance any given feature descriptor's performance. The image transforms that are available to the GA for learning such a composition are defined by the user. For the sake of article compactness, the full set of image transforms (20 in total) available to the GA have been omitted (they can be found in [1]). iECO has a major advantage over related feature learning works such as *convolutional neural networks* (CNNs) in the respect that it includes a set of heterogeneous image transformations, but is not a black box, i.e., each feature composition (individual's chromosome) can be "opened up" and studied. Individuals comprising the GA's population are allowed to have chromosomes of varying length. However, the maximum length has been limited to eight genes, as in both our previous work [1] and Lillywhite's original ECO work [3]. While limiting chromosome length is not required, it is most often done to reduce computation (run-time and learning) time. A chromosome is the segment of genes, i.e., series of image transforms (of which ordering matters), that represents a potential solution to the optimization task.

A key aspect of the iECO framework is that it is feature descriptor dependent. That is, if using multiple feature descriptors, iECO must be employed for each descriptor independently. In [1], it was shown that a descriptor's performance is degraded, sometimes dramatically, when interchanged with other descriptors' learned iECO pipeline. Thus, for each feature descriptor used, a unique iECO pipeline is learned. Herein, four feature descriptors are implemented; the (1) *histogram of oriented gradients* (HOG)[4], (2) *local binary pattern* (LBP)[5], (3) *edge histogram descriptor* (EHD)[6], and (4) *statistical descriptor* (SD)[1]. We note that local skewness and the difference between (each) local and global skewness has been added to the SD.

A common shortcoming of GAs is their lack of population diversity. The diversity promoting constraints introduced in [1] address this issue and ensure a more thorough search of the solution space. Let $\mathbf{\Theta} = [\theta_1, \theta_2, ..., \theta_n]$ denote n constraints, where n is the maximum allowed chromosome length. The value at θ_i defines the maximum percentage of gene overlap at the i^{th} gene of the chromosome allowed for all individuals in the population whose genes 1 through $i-1$ overlap as well (i.e., are the same). Thus, through the defining of $\mathbf{\Theta}$, great control and assurance over the amount of diversity present in an evolutionary algorithm's population is obtained.

In [2], importance maps for per-pixel weighting are derived from *differential morphological profiles* (DMPs). The DMP can exploit contrast edges between objects and their surrounding context to extract said objects. Using geodesic morphological reconstruction, the DMP extracts objects that are lighter (opening) and darker (closing) than their surrounding context. First, we apply a median filter for denoising and edge preservation. The DMP produces a set of scale-attributed responses using a geodesic disk of size r_m, where $m \in M$ defines the scales of a morphological structuring element. Herein, geodesic disks of radii 1, 3, 5, 7, 9, and 11 m are used. By computing the piece-wise differential, i.e., response at scale r_m minus r_{m-1}, we find objects that survive up to some scale of SE, then are obliterated in a subsequent scale. Thus, levels in the DMP are the set of objects extracted during a particular geodesic scale transition. After computing the DMP, it is fused into a soft segmentation (per pixel confidence) of objects in a ROI based on the *Choquet integral* (CI). Finally, the soft segmentation (called a importance map hereafter) is obtained by processing the fused DMP image with morphological 3 m radial dilation and object reconstruction, which further enhances the soft segmentation and isolates the object of interest. We then compute the eigenvalues and eigenvectors of the soft segmentation in order to perform

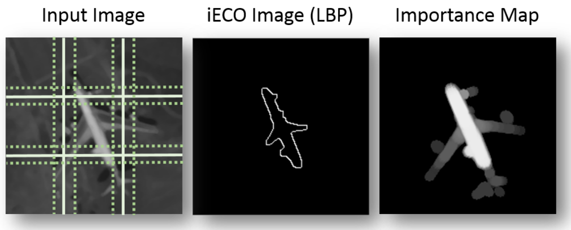

Fig. 3. Example of input image, an iECO output, and its corresponding importance map. The dashed lines provide a visual illustration of the partial overlap across cells.

a rotational alignment of the original image and the importance map. Due to ambiguity in the direction of the major eigenvector, we perform both possible rotations and use both in training. For further details on importance map generation, refer to [2]. Per-pixel features, such as the HOG, LBP, *Haar-like features* (HLF), and the *invariant object moments* (IOM), make use of the importance map by using it when calculating a histogram of these features. Instead of adding these features to a histogram as they have always been added, one simply multiplies their corresponding $[0, 1]$ importance map value by the amount that would otherwise be added to the descriptor.

This article is interested in studying the applicability of iECO for remotely sensed imagery. However, we are also interested in the exploration of using the iECO framework in conjunction with soft features. To some extent, these two methods are pursuing similar end-goals. That is, soft features are employed to extract an object of interest's defining characteristics and to "ignore" background, i.e., non-target information. Similarly, the iECO framework seeks to find an optimal composition of image transforms such that a given feature descriptor can best extract information for discriminating target from non-target. Therefore, it is advantageous to explore the combination of these two theories in order to learn what features are best in a context independent fashion. It is relatively simple to combine these two techniques. As we calculate the histogram features for iECO, we simply use the corresponding $[0, 1]$ importance map weight in the update rule (multiplication of the standard update by the importance map weight).

In total, there are four approaches to object recognition in satellite imagery investigated; raw, soft (importance map weighted) features, iECO, and the combination of soft features with iECO. For illustrative purposes, Figure 3 shows an example chip from our database, a corresponding iECO feature, and the derived importance map.

3. EXPERIMENTAL FINDINGS

Experiments are performed based on four-fold (scene-based) *cross-validation* (CV). That is, for each of the four scenes, one is designated as the testing scene and the remaining three scenes are used for training. Furthermore, each scene corresponds to a data collection from different times of the year. The data consists of three classes of target objects: commercial jets, helicopters and military cargo. Depending on the scene, the number of target objects for each class varies. In addition, there are (pseudo)random image chips in which no target class instance is present. A one versus all classification approach is employed. Results are presented using a *target recognition rate* (TAR) and *false alarm rate* (FAR) based table summary of our findings.

To begin, we report and discuss the results under the most optimistic scenario: re-substitution. Herein, numerous kernels and parameters were tried; however, the polynomial kernel with degree 3 was chosen due to it having the best performance for this data set and re-substitution experiments. For compactness, the average TARs and FARs for each method across all scenes and objects is reported in Table 1.

Table 1. Averaged TARs and FARs for re-substitution.

	TAR	FAR
Raw	0.984	0.000
Soft	0.769	0.000
iECO	1.000	0.000
iECO+Soft	0.844	0.001

From Table 1, we see that iECO performs the best, achieving perfect classification for our data. Comparatively, raw features performs very well with a slight reduction in TAR, but no decrease in FAR, and soft features has a surprisingly large drop in TAR (again, no drop in FAR). The effects of soft features is seen in the performance of iECO+soft as TAR is reduced (or, conversely, we see iECO brings the performance of soft features up). However, due to the drop in TAR performance between iECO+soft and iECO individually, we realize that soft features (i.e., the importance map) has the possibility to degrade/hinder the information that iECO has learned to queue on. At the same time, iECO has the ability to gather such discriminative information that it is able to still improve the performance of soft features (used by themselves) even with its information being, at times, degraded by the importance map. This indicates that iECO features are very powerful and robust.

Next, an analysis of the four-fold CV results is provided. For these experiments, a number of different kernels were explored (e.g., *radial basis function* (RBF), polynomial, etc.) and it was found that a linear *support vector machine* (SVM) performed the best for this data. Overall, the combination of iECO and soft feature extraction performed the best. Again, for compactness, we only report the average TARs and FARs for each method across all scenes and objects, and this is given in Table 2.

From the initial results presented in Table 2, we see that the combination of iECO and soft features results in the

Table 2. Averaged TARs and FARs for four-fold CV.

	TAR	FAR
Raw	0.815	0.045
Soft	0.590	0.018
iECO	0.802	0.019
iECO+Soft	0.893	0.035

best TAR performance, although it comes at the expense of a (slightly) higher FAR than both soft features and iECO individually. One possible reason for iECO+soft features' increased FAR is iECO could be cueing on information that, at times, gets diminished by the importance map. However, to put the TARs and FARs into some perspective, we are talking a difference in an average of 1.5 more false alarms versus an average of 9 additional positive detections (in reference to iECO + soft and iECO individually). In all three cases, soft, iECO, and iECO+soft, the FAR is less than raw features. Interestingly, focusing on just TAR performance, soft features perform the worst, which supports the re-substitution results (Table 1). The exact reason(s) for such inferior performance by soft features is subject for further research. However, we do want to mention some of our initial thoughts on the performance of soft features in this work. Soft features are not a simple straight-forward technique. They require a number of pre-processing steps and parameters for operation, all of which impact the final importance map that is produced and used for feature weighting. Additionally, we may need to reassess our approach to deriving the final fused DMP image (which is our importance map). We have a number of DMPs that are primed to extract objects at different scales, i.e., one DMP is better for detecting small objects (e.g., helicopter blades), while another is better at detecting large objects (e.g., fuselage on a commercial jet). Rather than fusing these to generate a final importance map, it may be of more benefit to keep these separate, and design a classifier for each object type in which each object's classifier output is fused to arrive at a final decision. Finally, iECO alone performs well. While iECO does not have as high TAR as raw features, iECO exhibits a better FAR (false alarms are reduced by more than half).

4. CONCLUSION

In conclusion, we found that iECO is able to learn very robust and discriminative features. Additionally, results reported in Table 2 indicate that there is indeed potential in the combination of iECO and soft features; however, this needs further research to better understand the complicated nature of these two methods being intertwined. For example, from the re-substitution experiments, soft features hurt the performance of iECO when combined, yet improved iECO's performance on the four-fold CV experiments. One possible conclusion could be that combining iECO and soft features results in a more generalized framework, but with the expectation that iECO could be hindered by the importance map at times while iECO very reliably improves the performance of soft features alone. This of course requires a much more thorough investigation into the combining of these two techniques before any such conclusions could be made. Furthermore, the performance of soft features is surprising. It is still our belief that soft features are a valuable technique, but its parameters and fusion may need to be revisited. Meaning, were the importance maps derived properly? It is plausible that different morphological operators and/or parameter selection could drastically improve the quality of importance maps generated. Future work will consist of further exploring how to better combine these two techniques and improving the fusion of the information across the iECO population.

5. ACKNOWLEDGMENTS

This work is partially funded by the Army Research Office grants numbered W911NF-14-1-0114 and 57940-EV to support the U.S. Army RDECOM CERDEC NVESD.

6. REFERENCES

[1] Stanton R Price, Derek T Anderson, and Robert H Luke, "An improved evolution-constructed (ieco) features framework," in *Computation Intelligence, 2014 IEEE Symposium Series on*. IEEE, 2014.

[2] Grant J Scott and Derek T Anderson, "Importance-weighted multi-scale texture and shape descriptor for object recognition in satellite imagery," in *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*. IEEE, 2012, pp. 79–82.

[3] Kirt Lillywhite, Dah-Jye Lee, Beau Tippetts, and James Archibald, "A feature construction method for general object recognition," *Pattern Recognition*, vol. 46, no. 12, pp. 3300–3314, 2013.

[4] David G Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[5] Timo Ojala, Matti Pietikäinen, and David Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, pp. 51–59, 1996.

[6] K Stone, JM Keller, DT Anderson, and DB Barclay, "An automatic detection system for buried explosive hazards in fl-lwir and fl-gpr data," in *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics, 2012, pp. 83571E–83571E.