AUTOMATIC OBJECT EXTRACTION FROM FULL DIFFERENTIAL MORPHOLOGICAL PROFILE IN URBAN IMAGERY FOR EFFICIENT OBJECT INDEXING AND RETRIEVALS

C.R. Shyu[a,b], G. Scott[a,b], M. Klaric[a,b], C. H. Davis[a,c], K. Palaniappan[a,b]
[a]Center for Geospatial Intelligence
[b]Department of Computer Science
[c]Department of Electrical and Computer Engineering
University of Missouri, Columbia, MO 65211-2060, USA
{shyuc, grantscott, mnkkrc, davisch, palaniappank}@missouri.edu

KEY WORDS: Urban, Object, Extraction, Segmentation, Multispectral, Retrieval

ABSTRACT

The differential morphological profile (DMP) can be used for automated extraction of multi-scale urban features, such as buildings, shadows, roads, and other man-made objects. However, characterization of urban features using the DMP is complicated by the fact that some objects will have response at multiple-scales within the DMP. This makes robust and efficient object indexing and retrieval difficult for large-scale remote sensing image databases utilized in the defense and intelligence communities. To address this issue, in this paper we present a novel approach called Multi-scale Extraction of Morphological Objects (MEMO),which is fully automatic and unsupervised. MEMO contains two processing modules for identifying urban objects: (1) Top-down object fusion: multi-scale objects from both morphological closing and opening profiles with certain topological relationships (TR), such as overlap, equal, and inside, will be selected for candidate objects and placed in the candidate pool. (2) Knowledge-based filtering: objects of the DMP are refined and filtered using information present in the original panchromatic image, spectral information of the scene, and the processed DMP. For example, areas of vegetative land cover are filtered out reducing the false labelling of tree clusters and fields candidate objects. Additionally, size and shape analysis of candidate objects can further eliminate possible false object extraction. The efficiency of our algorithm makes it applicable to large-scale object indexing and retrieval.

1 INTRODUCTION

Recent advances in the quality of satellite imagery and the desire to analyze this data has spurred the development of new image processing techniques for object extraction. There are many applications for which it is useful to be able to efficiently extract individual objects from a scene for the purpose of spatial analysis and object retrievals from large-scale image databases. Items of interest include man-made structures such as buildings and roads. The techniques used for such analysis employ various methods to appropriately segment an image to highlight the relevant information of the scene.

Prominent techniques in object extraction from remote sensing images include the approaches presented in (Salembier and Pardas, 1994, Stilla et al., 2003, Shackelford et al., 2004, O'Callaghan and Bull, 2005). These approaches roughly fall into two categories: edge-detection approaches or region analysis approaches (Benediktsson et al., 2003). Among the region analysis approaches, the Differential Morphological Profile (DMP) (Persaresi and Benediktsson, 2001) draws much attention recently. DMP is a multi-scale image processing algorithm that employs a combination of morphological operators and derivatives of the resulting morphological profile. A multi-scale approach is required for two reasons. For the general case of object extraction, it is impossible to assume that all objects will always be of a uniform size in a given image; thus, structuring elements of multiple scales are a necessity. Additionally, even if this assumption was possible, a multi-scale technique also allows for the identification of object substructure.

By varying the size of the structuring element used in the morphological operations, different regions corresponding to candidate objects will provide different levels of response. Objects with sizes varying from small to large will show an increased level of response as the size of the structuring element is changed correspondingly. The results of applying the DMP algorithm can then be analyzed to determine the position of maximum response for each pixel in the image. This can be used to gauge the approximate size of an object at that location and its contrast with the surrounding area (Soille and Pesaresi, 2002, Shackelford et al., 2004). Alternatively each pixel in the DMP can be treated as a multi-dimensional vector for the purposes of feature extraction. Techniques such as discriminant analysis feature extraction (DAFE) and decision boundary feature extraction (DBFE) were proposed in (Benediktsson et al., 2003). Classification techniques such as DBFE require that a neural network be trained to correctly identify object pixels of an image with high degree of accuracy. However, this type of approach may require a neural network for each scene. In addition, pixel-based classification is impractical for object-based information indexing and retrieval.

The approach presented in this paper is a fully automated technique for rapid object extraction. While previous pixel-based methods are not practical for some information retrieval applications, a truly object-based approach generates information which can be used in systems that allow for queries based on object level characteristics. Additionally, our approach is a general technique that can be applied to scenes from many different images. The experiments presented show the results from applying our algorithm to five scenes. Furthermore, the method presented is highly

efficient; automated object extraction occurs in a matter of minutes. In addition, our approach does not require the training phase that is a requisite of other techniques.

This paper is organized as follows. Section 2 describes the algorithms used in our approach. A brief overview of morphology principles is presented before presenting our method. Section 3 discusses the evaluation techniques used. The experimental results presented are analyzed with the metrics presented in (Mariano et al., 2002). Although no single measure can capture the complete performance of an object extraction algorithm, the metrics presented can allow for a comparison to a ground-truth set of objects identified in a scene. Next, Section 4 details the results from applying our algorithm to five scenes from two high-resolution satellite images. Finally, Section 5 concludes with a discussion of potential applications of our method and ideas for future work.

2 ALGORITHMS

The core principal of the MEMO algorithm is the top-down fusion of a differential morphological profile (DMP) (Persaresi and Benediktsson, 2001) response into an object set. The DMP is generated using a set of increasing disk structuring elements (SE) and grayscale morphological reconstruction operators. Traditionally, the DMP is considered as a vector response of each pixel to the DMP algorithm. The resulting vector depicts the SE-size ordered differential response of opening by construction followed by the reversed ordered differential response to closing by reconstruction. The response of the DMP in the levels of a larger SE reflects the presence of larger objects, and the response in levels corresponding to a smaller SE indicates smaller objects.

For the purpose of MEMO, the DMP is divided into two profiles, *closing morphological profile* and *opening morphological profile*. Many other approaches utilize a pixel classification approach, and then extract objects using the classification of a region of pixels (Benediktsson et al., 2003). In contrast, MEMO functions at the object level by analyzing each level of both the opening and closing profiles separately. The final object detection is the result of top-down object fusion in each of the profiles combined into a single object response set. The pseudo code of MEMO object extraction is listed in Algorithm 1.

The input of the MEMO algorithm contains a panchromatic, remotely sensed urban image; near infrared spectral response; and red spectral response of a scene. The algorithm applies a series of set, morphological, and image processing/computer vision operations. The first two steps of MEMO include generating a Sobel edge detection response and a normalized difference vegetation index (NDVI) response. Next, the DMP of the panchromatic image is generated and divided into the closing and opening differential profiles, DMP_C and DMP_O, respectively. Each level of the DMP forms a complete grayscale image. Starting with the largest SE-size level, each level of both profiles is analyzed for candidate objects in a top-down manner.

Algorithm 1 MEMO Object Extraction

1: SOBEL ← Generate a Sobel edge detector response map from PAN
2: NDVI ← Generate a NDVI response map from NIR and Red-channel.
3: $\{\text{DMP}_C,\text{DMP}_O\}$ ← Generate DMP from PAN, organized into opening and closing profiles.
4: **for all** $_p \in \{\text{C,O}\}$ **do**
5: **for all** l=max_level to min_level **do**
6: $\text{DMP}_{p,l}$ ← Vege_Filter($\text{DMP}_{p,l}$, NDVI, threshold)
7: $\text{DMP}_{p,l}$ ← Max($\text{DMP}_{p,l}$, SOBEL)
8: $\{\text{OBJECTS}_{p,l}\}$ ← GL_Connected_Components($\text{DMP}_{p,l}$)
9: Merge each $\{\text{OBJECTS}_{p,l}\}$ into MEMO_p
10: **end for**
11: **end for**
12: MEMO ← Merge the opening, $\{\text{MEMO}_O\}$, and closing, $\{\text{MEMO}_C\}$

In Algorithm 1, we use p to represent either opening or closing profile and l for each profile level ranging from max_level to min_level. We depict level l of profile p with the notation $\text{DMP}_{p,l}$. Vege_Filter() in step 6 filters $\text{DMP}_{p,l}$ by removing all pixels that have an NDVI response greater than a threshold. In our work, we use 0.2 for this threshold. The benefit of this operation is the removal of most vegetation in the image seen, such as grass fields, large trees and tree clusters. Each $\text{DMP}_{p,l}$ is then fused to the Sobel response using point-wise maximum function, Max(), which takes the higher value of each pixel position from two images. This is motivated by the benefit it provides to the decomposed gray level connected components, Algorithm 2, by separating close-together objects with identical DMP response.

After the $\text{DMP}_{p,l}$ is refined in steps 6 and 7, the challenge is to find the connected components with a grayscale approach. When applying a traditional component labelling, the grayscale image may result in thousands of labeled classes in the smaller SE levels of the DMP. The amount of computation required to resolve label equivalency classes hinders applicability to real-time object extraction for object retrievals. Many approaches to grayscale component labelling have been explored in literature (Gonzalez and Woods, 2002; Brag-Neto and Goutsias, 2004), however we chose to develop an algorithm which exploits properties inherent in a single profile level. It can be observed that given a particular level of the DMP, the number of grayscale values in the response is usually quite small for an urban image, always less than 50 distinct grayscale values in our testbed scenes. This allows us to decompose a DMP level, $\text{DMP}_{p,l}$, into a set of binary images. Each image, $\text{DMP}_{p,l,i}$, is composed of all pixels of intensity i from $\text{DMP}_{p,l}$.

Algorithm 2 lists the pseudo code of our specialized grayscale component labelling. The algorithm first initializes

Algorithm 2 GL_Connected_Components

1: Intialize $\{\text{OBJECTS}_{p,l}\}$
2: Histo[] $\leftarrow$ Histogram of $\text{DMP}_{p,l}$
3: **for all** $i > 0$ and Histogram$[i] > 0$ **do**
4: $\text{DMP}_{p,l,i} \leftarrow$ set of pixels from $\text{DMP}_{p,l}$ of intensity i.
5: $\text{DMP}_{p,l,i} \leftarrow$ Morphological Opening of $\text{DMP}_{p,l,i}$ by SE_l.
6: $\{\text{OBJECTS}_{p,l,i}\} \leftarrow$ Binary Connected Components of $\text{DMP}_{p,l,i}$
7: Filter $\{\text{OBJECTS}_{p,l}\}$ using size and shape measures
8: $\{\text{OBJECTS}_{p,l}\} \leftarrow$ Max($\{\text{OBJECTS}_{p,l}\}$, $\{\text{OBJECTS}_{p,l,i}\}$)
9: **end for**
10: Return $\{\text{OBJECTS}_{p,l}\}$

an empty image. As a preprocessing step to this algorithm, the $\text{DMP}_{p,l,i}$ is morphologically opened with a level-dependent structuring element, SE_l. This effectively eliminates thin connections between components. A traditional binary component labelling algorithm is run against each $\text{DMP}_{p,l,i}$, extracting objects that share a grayscale value. The elements of the object-set, $\{\text{OBJECTS}_{p,l,i}\}$, generated from each relevant gray-level, are evaluated based on object size and shape measures to filter out unsuitable objects. For example, objects with unreasonably large scene coverage are excluded. To filter highly irregularly shaped objects, each detected object is compared to its bounding box and those with low density are also excluded. The object-set, $\{\text{OBJECTS}_{p,l}\}$, for $\text{DMP}_{p,l}$ is accumulated from the satisfactory objects of each $\{\text{OBJECTS}_{p,l,i}\}$. We know any two object sets from the same $\text{DMP}_{p,l}$ contain no overlapping objects because of the following property: $OBJECTS_{p,l,i} \cap OBJECTS_{p,l,j} = \emptyset, \forall i \neq j$. This final object-set is returned to Algorithm 1.

Each profile level, *l*, merges the current profile's detected objects, MEMO_p, with any previously detected objects (i.e. larger scale objects detected with larger SE). The final step is to combine the objects detected in the opening morphological profile, MEMO_O, with the objects detected in the closing morphological profile, MEMO_C, into a final object set. Our output preserves DMP level information by using increasing grayscale values for detected objects from decreasing profile levels. Figure 1 provides an example of the stages of MEMO object detection. Objects detected at lower levels are represented with higher grayscale values in the MEMO output and may be superimposed onto objects detected at higher levels.

3 EVALUATION METHODS

The methods used for evaluating MEMO quantify several different facets of the building detection algorithm. The quality of the results produced by our system is measured by comparing the detected objects to a set of ground-truth objects. These ground-truth images were generated by a human analyst manually identifying the polygon which surrounds target objects.

There exists no single measure which can be used to evaluate all aspects of the performance of object detection. The algorithms used for object detection evaluation fall into two general categories as described in (Mariano et al., 2002). The proposed metrics are either pixel-based or object-based measures. Additionally, the methods used either focus on determining the rate of true-positive or false-positive results. A pixel or an object is identified as valid if it was correctly identified as belonging to a object in the ground-truth image. Conversely a false-positive response occurs when a pixel or object was incorrectly detected.

Several of the measures used for evaluation rely upon the following two formulas. The set U_G and U_D are the union of all ground-truth and detected objects, respectively.

$$U_G = \bigcup_{i=1}^{N_G} G_i \qquad U_D = \bigcup_{i=1}^{N_D} D_i \tag{1}$$

Two pixel-based evaluation methods were used. The first measure is called area-based recall. This metric determines how well the detected pixels intersect the manually identified pixels of the ground-truth image.

$$PixelRecall = \begin{cases} undefined & if\ U_G = \emptyset \\ \frac{|U_D \cap U_G|}{|U_G|} & otherwise \end{cases} \tag{2}$$

Area-based precision is the second pixel-based metric. This process measures how well the the pixel regions of the detected and ground-truth images overlap; it captures the tendency of the object extraction algorithm to introduce false-positives into the results.

$$PixelPrecision = \begin{cases} undefined & if\ U_D = \emptyset \\ 1 - \frac{|U_D \cap \overline{U_G}|}{|U_D|} & otherwise \end{cases} \tag{3}$$

Additionally, several object-based metrics were considered. The first of these measures is fragmentation. This metric is formulated in such a way that an algorithm is penalized for incorrectly splitting an object into multiple objects during detection; this results in multiple objects which overlap a single ground-truth object.

$$GTfrag(G_i) = \begin{cases} undefined & if\ N_{D \cap G_i} = 0 \\ \frac{1}{1+log(N_{D \cap G_i})} & otherwise \end{cases} \tag{4}$$

Conversely a second fragmentation measure calculates the frequency with which an algorithm produces objects which overlap multiple ground-truth objects. We found this measure is particularly useful for determining how well the algorithm can distinguish adjacent objects with weak boundaries.

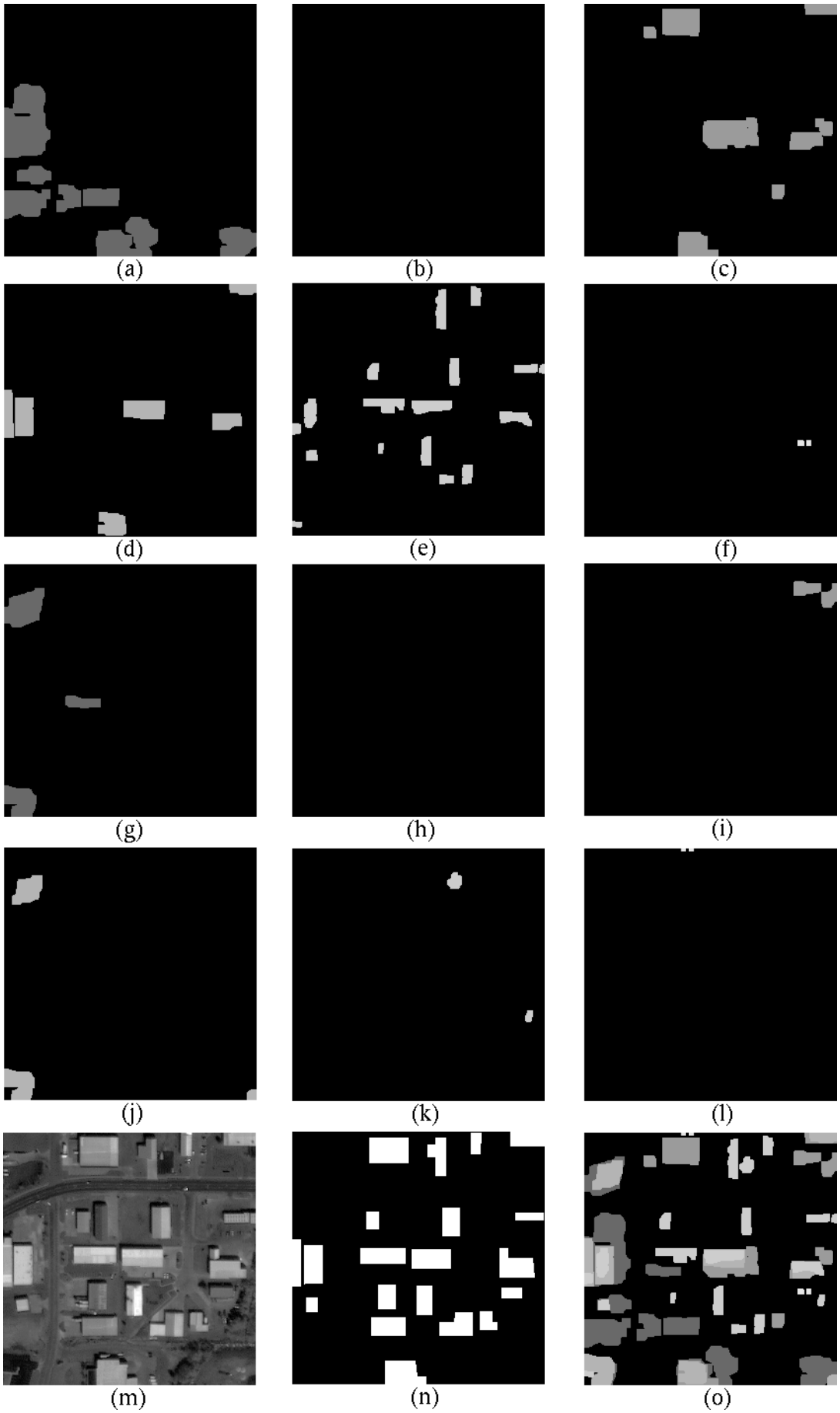

Figure 1: MEMO Step-by-step intermediate results for scene Columbia 3:(a)-(f) Objects from opening levels six through one sequentially. (g)-(l) Objects from closing levels six through one sequentially. (m) Original panchromatic scene. (n) Analyst identified objects. (o) MEMO identified objects

Scene	Recall	Precision
Columbia 1	52%	49%
Columbia 2	58%	64%
Columbia 3	76%	42%
Springfield 1	64%	42%
Springfield 2	61%	43%
Average	62%	48%

Table 1: Pixel-based Measures

$$DTfrag(D_i) = \begin{cases} undefined & if\ N_{G \cap D_i} = 0 \\ \frac{1}{1+log(N_{G \cap D_i})} & otherwise \end{cases} \quad (5)$$

Another object-based measure computes the percentage of each ground-truth object that is correctly identified; this is called the object area recall. When an absolute threshold, ϵ, is applied to this value, a list of successfully detected ground-truth objects can be created. In our experiments, we use $\epsilon = 0.5$.

$$ObjRecall(G_i) = \frac{|G_i \cap U_D|}{|G_i|} \quad (6)$$

$$ObjDetected(G_i) = \begin{cases} 1 & if\ \frac{|G_i \cap U_D|}{|G_i|} > \epsilon \\ 0 & otherwise \end{cases} \quad (7)$$

Similarly, a measure of the detected object area precision is used to calculate what portion of a detected object intersects the ground-truth region. This ratio is used to analyze the number of false-positive objects generated by the algorithm. Applying a threshold, ϵ, to this value yields a list of those objects which were detected as matching the ground-truth objects. Again, we used $\epsilon = 0.5$ for our experiments.

$$DetectedObjPrecision(D_i) = \frac{|D_i \cap U_G|}{|D_i|} \quad (8)$$

$$DetectedObjAccepted(D_i) = \begin{cases} 1 & if\ \frac{|D_i \cap U_G|}{|D_i|} > \epsilon \\ 0 & otherwise \end{cases} \quad (9)$$

4 RESULTS

The MEMO building extraction algorithm was applied to several scenes from 1-m panchromatic IKONOS images of downtown Columbia and Springfield, Missouri. Each scene is processed to generate a DMP, which is then analyzed in a top-down fashion using Algorithm 1. In each profile, candidate objects of the level corresponding to the largest SE are continuously refined by iteratively merging objects found in successively smaller levels of the profile.

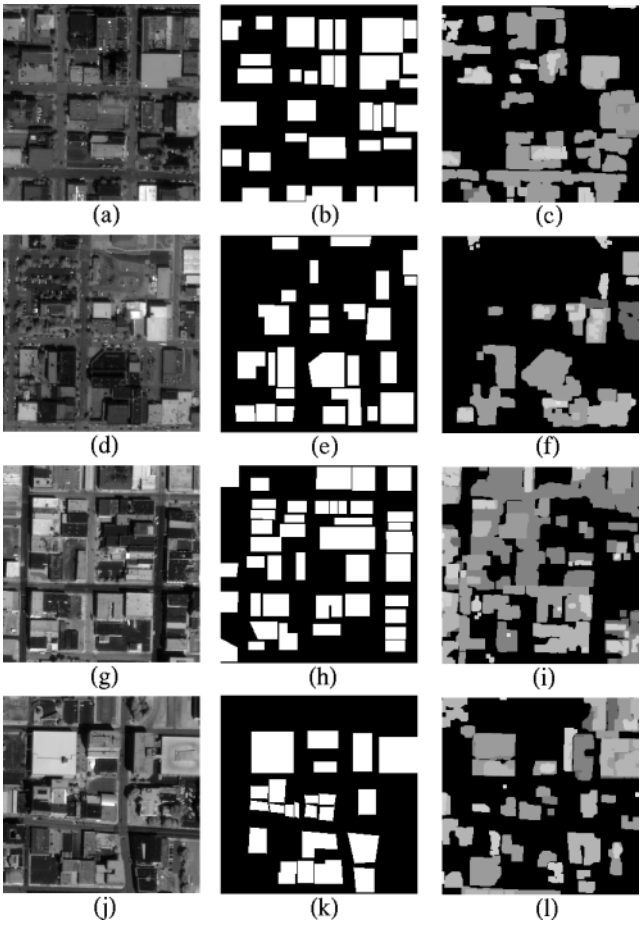

Figure 2: (a) Original panchromatic image, (b) human identified buildings, and (c) MEMO identified buildings from scene Columbia 1. (d)-(f) scene Columbia 2, (g)-(i) scene Springfield 1, (j)-(l) scene Springfield 2

The entire object extraction process, including the generation of the DMP, occurs in under 3 minutes on a modern Pentium 4 machine.

As shown in Table 2, on average our algorithm correctly extracts nearly 66% of the objects in the five scenes using Eq. 7. The example depicted in Figure 1 demonstrates a sample execution of this algorithm in which 18 of 20 buildings were successfully extracted. This example also shows the algorithm's low tendency to fragment buildings into two detected objects. Ground-truth fragmentation–the measure which corresponds to this tendency–is calculated at 95% using Eq. 4 for this example. The average value across all tested scenes is 83%.

However, the MEMO algorithm does have a tendency to extract erroneous objects in some situations. In the example shown in Figure 1, only 56 percent of the objects detected correspond to ground-truth objects by Eq. 9; this is slightly above the average for our tests. Additionally, some of these situations correspond to cases when many detected objects become incorrectly merged into one. Using Eq. 5, the metric for the fragmentation of detected objects evaluates to 67% for the scene depicted in Figure 1, which is also the average value of this measure.

Scene	Object Area Recall	Ground-truth Objects Detected	Object Area Precision	Detected Objects Accepted	Fragmentation of Ground-truth Objects	Fragmentation of Detected Objects
Columbia 1	54%	63%	48%	50%	81%	70%
Columbia 2	51%	54%	51%	64%	71%	56%
Columbia 3	73%	90%	52%	56%	95%	67%
Springfield 1	69%	72%	51%	61%	90%	80%
Springfield 2	49%	52%	41%	46%	77%	62%
Average	59%	66%	48%	55%	83%	67%

Table 2: Object-based Measures

5 CONCLUSIONS

Our future efforts will further refine the algorithm. Currently the object extraction employs heuristic analysis of object size and shape features to determine candidate objects in each level. However, due to the variability of buildings in urban remote sensing images, current heuristics must be complimented with additional methods. Other avenues to be explored would be additional image processing and computer vision techniques. A noticeable effect of the DMP is that larger SE-sizes tend to extract large blobs that are composed of multiple smaller objects and sometimes relevant objects with non relevant objects. An example of this is seen by examining the MEMO output in Figure 2 (g)-(i), where some roads form blobs that include buildings and parking lots of similar intensity.

MEMO is a key part of our recently developed Geospatial Information Retrieval and Indexing System (GeoIRIS) for urban object characterization and retrieval. Successful automatic object extraction algorithms coupled with other geographic data allow GeoIRIS to support the following queries for use by intelligence analysts: (1) Object query: find database objects similar to a given query object. (2) Nearest neighbor query: find objects spatially close to a given object. (3) Distance scan: find objects within a certain distance of a given object, possibly with direction constraints. (4) Object/area query: find objects similar to a given query object and its surrounding area.

ACKNOWLEDGMENTS

This project is currently supported by the National Geospatial-Intelligence Agency University Research Initiatives (NURI) under grant number HM1582-04-1-2028.

REFERENCES

Benediktsson, J. A., Pesaresi, M., and Arnason, K. (2003). Classification and feature extraction from remote sensing images from urban areas based on morphological transformations. *IEEE Transactions on Geoscience and Remote Sensing*, 41(9):1940–1949.

Brag-Neto, U. and Goutsias, J. (2004). Grayscale level connectivity: Theory and applications. *IEEE Transactions on Image Processing*, 13(12):1567–1580.

Gonzalez, R. C. and Woods, R. E. (2002). *Digital Image Processing*. Prentice Hall, Upper Saddle River, New Jersey.

Mariano, V. Y., Min, J., Park, J.-H., Kasturi, R., Mihalcik, D., Li, H., Doermann, D., and Drayer, T. (2002). Performance evaluation of object detection algorithms. pages 965–969. Proceedings. 16th International Conference on Pattern Recognition.

O'Callaghan, R. J. and Bull, D. R. (2005). Combined morphological-spectral unsupervised image segmentation. *IEEE Transactions on Image Processing*, 14(1):49–62.

Persaresi, M. and Benediktsson, J. A. (2001). A new approach for the morphological segmentation of high-resolution satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 39(2):309–320.

Salembier, P. and Pardas, M. (1994). Hierarchical morphological segmentation for image sequence coding. *IEEE Transactions on Image Processing*, 3(5):639–651.

Shackelford, A. K., Davis, C. H., and Wang, X. (2004). Automated 2-d building footprint extraction from high-resolution satellite multispectral imagery. volume 3, pages 1996–1999. Proceedings of International Geoscience and Remote Sensing Symposium.

Soille, P. and Pesaresi, M. (2002). Advances in mathematical morphology applied to geoscience and remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 40(9):2042–2055.

Stilla, U., Michaelsen, E., Soergel, U., and Schulz, K. Perceptual grouping of regular structures for automatic detection of man-made objects. volume 6, pages 3525-3527. Proceedings of International Geoscience and Remote Sensing Symposium.