Finding Areas of Motion Using Camera-Trap Photos

Sara Beery and Agnieszka Miguel, Ph.D.

Department of Electrical and Computer Engineering, Seattle University

Abstract

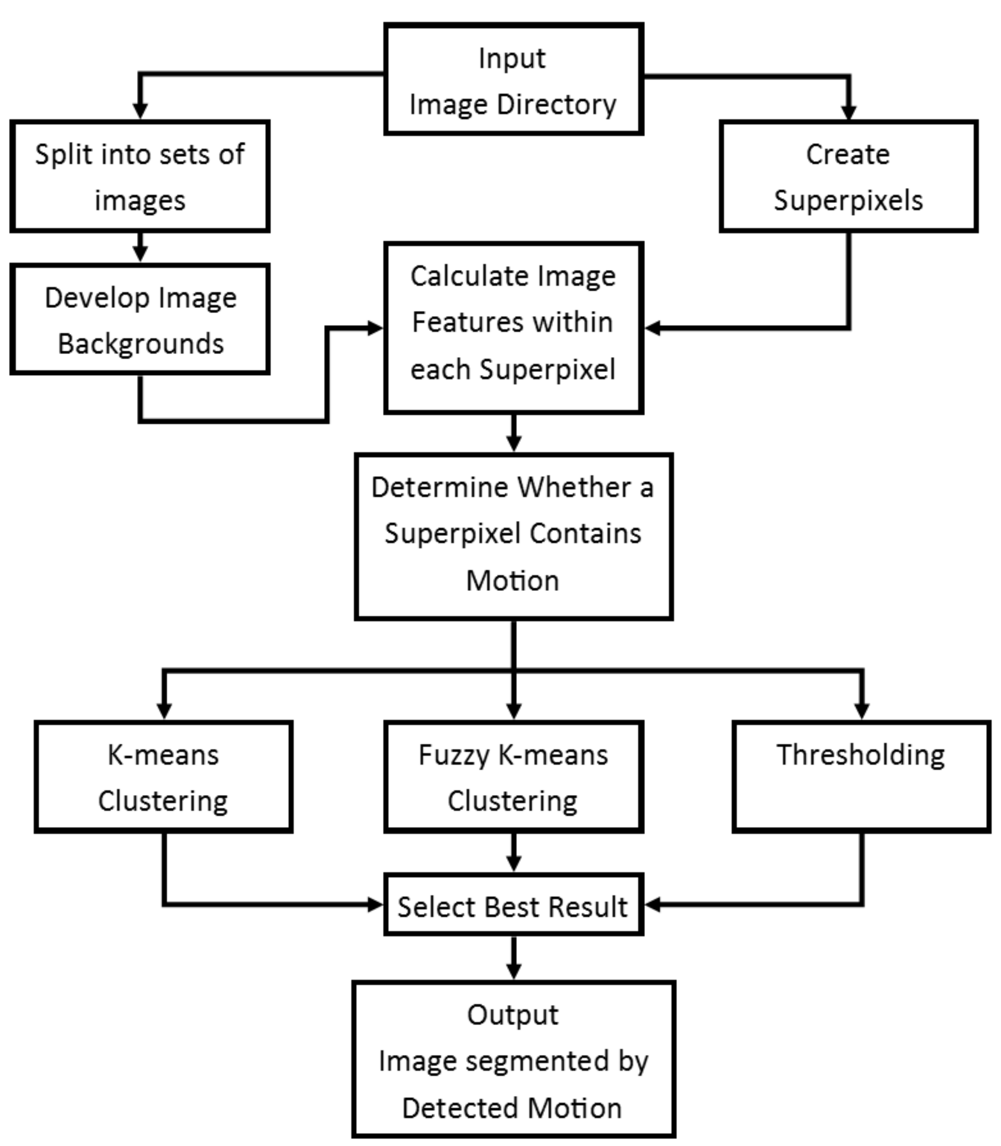
Camera trapping is used by conservation biologists to study snow leopards. In this research, we introduce techniques that find motion in camera trap images. Images are grouped into sets and a common background image is computed for each set. The background and superpixel-based features are then used to segment each image into objects that correspond to motion. The proposed methods are robust to changes in illumination due to time of day or the presence of camera flash..

Introduction/ Background

In our research we use images obtained from Panthera and Snow Leopard Trust. Each folder contains photographs taken from a certain location. A set of five images are captured when a moving heat source passes in front of the camera.

The camera can be triggered day or night by any type of animal, or by anything else moving that significantly shifts the temperature of the surrounding area.

The following diagram shows our proposed motion detection method.



Proposed Motion Detection Method

To locate the areas of motion in each image, we use concepts from background subtraction. Traditional background subtraction methods used to separate background from foreground in video sequences perform well when the background is static, for example indoors where the source of illumination does not vary. In the outdoor environments of camera trap images, changing weather (sun, rain, and wind) makes these methods very challenging to use. To account for the environment-related changes in the background, our first step is to sort all images from one location into **sets**. For each set, we then compute a common background image. Following, we use one of three methods to find the location of motion in each image (feature thresholding, k-means, and fuzzy k-means clustering).

Image Features

We use several types of superpixel-based texture features. Assume that image $I_i(x,y)$ has been segmented into *N* superpixels. In this work, *N*=10000. Subscript *i* indicates that image $I_i(x,y)$ is the *i*th image in set *k*. Its corresponding background image is $B_k(x,y)$. Let $S^n(I_i)$ be the set of all pixel coordinates that belong to the nth superpixel computed for image $I_i(x,y)$. We then define $I^n_i(x,y)$ as the segment of image $I_i(x,y)$ that corresponds to the *n*th superpixel:

$$yI_i^n(x,y) = \{I_i(x,y) : (x,y) \in S^n(I_i)\}$$

$$yI_i(x,y) = \bigcup_n I_i^n(x,y).$$

The **Mean of Differences** motion feature is computed as follows:

$$\text{MoD}\{I_i(x,y)\} = \bigcup_n \left\{ I_i^n(x,y) \leftarrow \frac{1}{|S^n(I_i)|} \times \sum_{(x,y)\in S^n(I_i)} |B_k(x,y) - I_i(x,y)| \right\},$$

The **Difference of Means** motion feature is computed as follows:

$$\text{DoM}\{I_i(x,y)\} = \bigcup_n \left\{ I_i^n(x,y) \leftarrow \frac{1}{|S^n(I_i)|} \times \left| \sum_{(x,y)\in S^n(I_i)} B_k(x,y) - \sum_{(x,y)\in S^n(I_i)} I_i(x,y) \right| \right\}$$

In addition to MoD and DoM, we also use **mean**, **range**, and **median** of superpixels as our features:

$$\text{Mean}\{I_i(x,y)\} = \bigcup_n \left\{ I_i^n(x,y) \leftarrow \frac{1}{|S^n(I_i)|} \times \sum_{(x,y)\in S^n(I_i)} I_i(x,y) \right\}$$

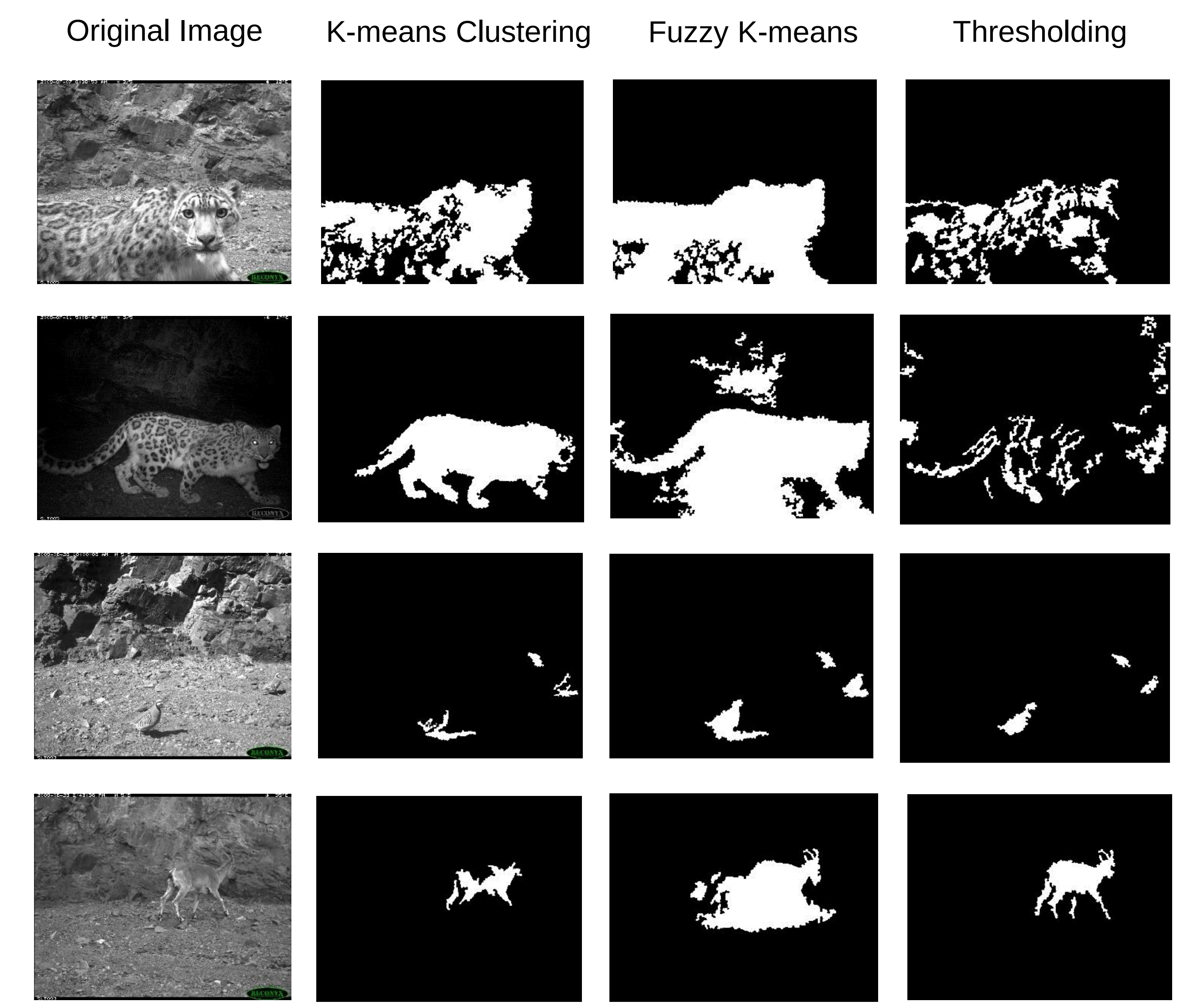
$$\text{Range}\{I_i(x,y)\} = \bigcup_n \left\{ I_i^n(x,y) \leftarrow \left(\max_{(x,y)\in S^n(I_i)} I_i(x,y) - \min_{(x,y)\in S^n(I_i)} I_i(x,y) \right) \right\}$$

$$\text{Median}\{I_i(x,y)\} = \bigcup_n \left\{ I_i^n(x,y) \leftarrow \text{median}_{(x,y)\in S^n(I_i)} I_i(x,y) \right\}$$

We then determine whether each superpixel corresponds to a region of motion based on these features using three separate methods: k-means clustering, fuzzy k-means clustering, and feature thresholding. We determine which method results should be selected by assuming the desired motion template should contain a small number of smooth objects and should be noise-free. Therefore, we count the number of objects in the template, analyze the smoothness of the chain code that describes their boundaries, and compute a frequency-based measure of noise in the template.

Results

We present examples of our promising results in the Figures below. For each original image, we apply the three proposed methods and show results in the corresponding rows of these two figures. No one method produces the best results for all types of images. This variability is present across all camera trap images and underscores the need to explore different segmentation methods for this particular application.



Conclusions and Future Work

The methods described here find motion in images from camera traps set to capture snow leopards. By using three different approaches, we produce results that are robust with respect to changes in illumination due to weather and time of day. This research will open the door to image classification based on type of animal sighted, and eventually to individual snow leopard recognition.

References

1. J. Tuszynski, Inscribed_Rectangle (Computer program), http://www.mathworks.com, July 2013
2. J. Motl, Local Image Thresholding (Computer program) http://www.mathworks.com, August 2013

The funding for this work has been provided by Center for Environmental Justice and Sustainability at Seattle University (SU), Clare Booth Luce Scholar Foundation, SU Summer Research Fellowship Program, and Francis P. Wood Chair in Electrical and Computer Engineering at SU. Thank you to Panthera, SLT, and SLCF Mongolia for data.

Seattle University Center for Environmental Justice and Sustainability Fellowship 2015-16