Map Seeking Circuits – A New Approach to an Old Problem

MIPS-SUBMISSION

Jerry G. Gregoire
Dr. Rob Maher
Montana State University

Although pattern recognition in both the visual and auditory domains is an extremely old problem with numerous intertwined research threads over the years, there remains a need for innovation in this field. Continued work on incremental improvements to the conventional techniques appears unlikely to produce the significant discoveries and breakthroughs that will lead to dramatic improvement in automatic detection and classification performance. Clearly, new ideas and techniques are desperately needed.

Arathorn (2002)[1] has recently proposed a new algorithm for pattern detection in images. The method, inspired by the neurobiology of the human visual system, performs an iterative comparison between a target pattern and a systematic set of allowable transformations applied to the observed image. The particular insight that makes this technique interesting is that the comparison process occurs on a single linear superposition of the allowable transformations, thereby eliminating the computational explosion that otherwise occurs when dealing with complicated images and patterns. The output of Arathorn's algorithm, as described below, is the set of transformations that map the target pattern to its location, translation, scaling, rotation, etc., within the observed image. Thus, Arathorn uses the term "map-seeking circuit" (MSC) to describe the computation.

Arathorn's insight is also particularly intriguing due to biological evidence that the visual system includes feedback paths in addition to the forward sensory neural pathways from the retina.

We have found that the unique features of the MSC algorithm for image processing show important promise in auditory scene analysis as well. Could this be the dramatic breakthrough that is needed in computational auditory scene analysis?

To find out, we are working on a set of experiments to extend and improve the MSC algorithm for use in auditory pattern detection and classification. A brief description of how a MSC works and how we are seeking to implement it in the audio realm follows.

A MSC has two paths, a forward path and a backward path, with a set number of transformation layers, As the audio mixture progresses through the forward path, it is transformed according to chosen parameters such as harmonic structure, pitch or duration. A similarity functions such as a dot product between the backward path and the layer transforms will accentuate the better transformations and deemphasize the other be adjusting a weight vector. After several iterations, the optimum transform for each layer will have a weight of one. All other transforms will have a weight below a given threshold and the search is complete. The resulting set of transforms now maps the chosen representation to the found object in the mixture. Reconstruction of the found object is a simple matter of applying the inverse transformations of each layer to the stored representation.

As an example, consider the visual example shown in Figure 1. The figure illustrates a MSC configured to find a square in a mixture containing both a square and a circle. The goal is to find the square and produce an image without the circle. The figure shows nine panels divided into two paths, the forward path is on the left and the backward path is on the right. This example assumes that if a square is in the mixture, a simple translation mapping will locate it. Therefore, it has two transformation layers, horizontal and vertical. Other variations such as rotation and scale are possible with additional transformation layers.

A mixture shown in panel 1 is presented to the MSC. Beginning on the upper left, the entire mixture is transformed by shifting it to the left. All the transforms are then added together as shown in panel 2. The second transform shifts the new image down as shown in panel 4. Using the similarity function to compare panel 4 to the stored representations, the square in panel 6 is chosen and propagates up the backward path. Again using the similarity function, panel 7 and panel 4 are compared to choose the favored layer 2 transform. In this case it is 'shift down one' and is shown in panel 5. On the backward path the stored target is shifted up by the inverse transform chosen in layer 2, 'shift up one'. This is shown in panel 8. The backward path is likewise compared to panel 2 to choose the favored layer 1 transform, 'shift left one', as shown in panel 3. One final inverse transform shifts the backward image right one to produce panel 9. As the incorrect transforms in the forward path are iteratively attenuated, the match with panel 6 is
strengthened. After several iterations the favored transformations are uniquely identified with the best target. At this point the target representation is now been mapped via the transformation set to the square in the original image.

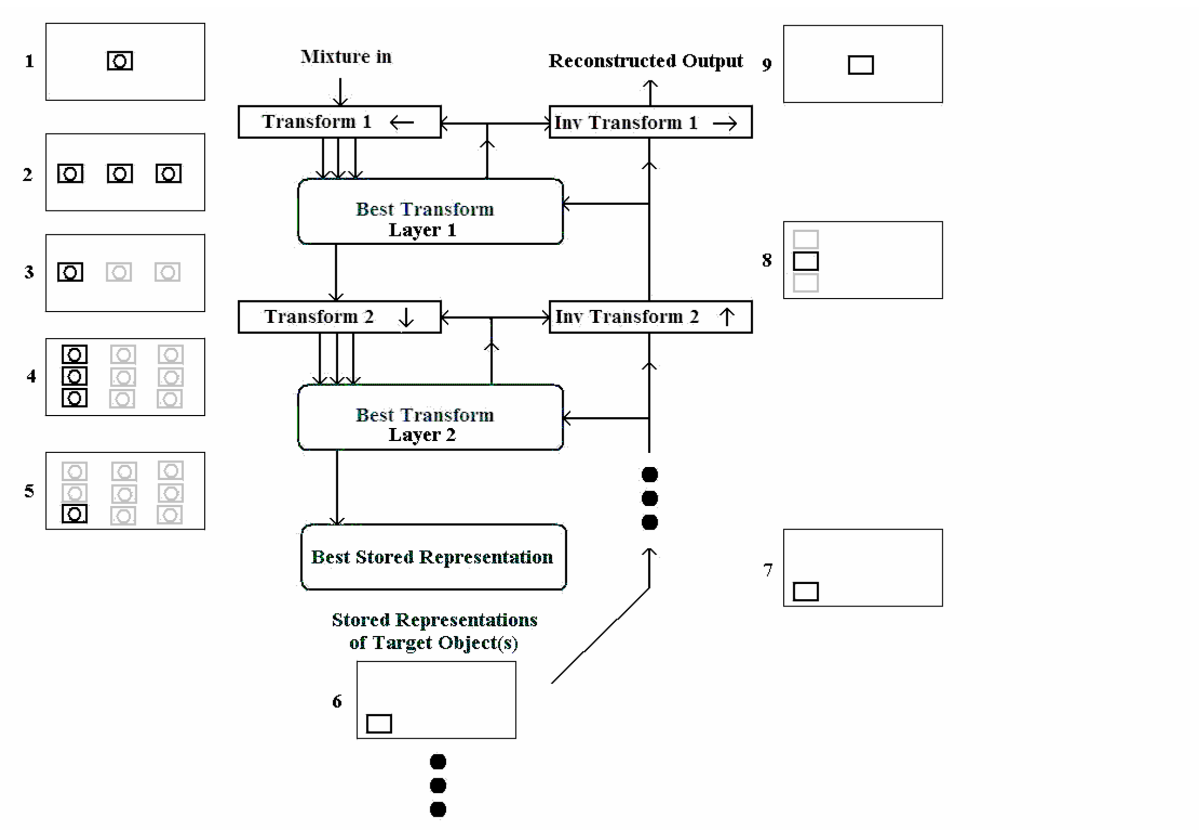

Figure 1: MSC Configured to Find a Square in a Mixture.

Choosing transforms that are both legitimate for audio processing and computationally efficient remain a challenge. We are currently seeking to implement audio based MSCs using both spectrograms and correlograms as the representational core. Our initial transform library includes pitch scaling, onset, and duration. Additional transforms may include among others, harmonic envelope, and modulation.

During the workshop we will discuss the details and theoretical underpinnings of MSC and the application to computational auditory scene analysis. We will also present our preliminary results and look forward to discussing options for other possible auditory representations and transforms.

---

[1] Arathorn (2002). Map Seeking Circuits: A Computational Mechanism for Biological and Machine Vision. Stanford, Stanford Press.