**Introduction**

In this paper we investigate a new method for segmentation of image sequences (such as in dynamic PET or fMRI). Our goal is to identify image regions in a dynamic study such that the image pixels within each region have nearly the same temporal activity functions. While many different methods exist for image segmentation, these methods are traditionally based on a Euclidean- or Mahalanobis-type metric [1, 2]. In the proposed approach we instead employ a metric based on the normalized cross-correlation, also known as the similarity, between two signals. Such a metric aims to capture the shape of the time functions rather than their overall amplitude.

We model the temporal activity in a dynamic image as a mixture of unknown basis functions. In the proposed method, these unknown basis functions, along with other unknown signal parameters, are found by maximum likelihood (ML) estimation using an expectation-maximization (EM) algorithm.

In this summary, we describe an initial study using a brain phantom model, which is used to simulate a dynamic PET study of [11 C] carfentanil binding to brain phantom model, which is used to simulate a PET or fMRI. Our goal is to identify image regions in a dynamic study such that the image pixels within each region have nearly the same temporal activity functions. For simplicity, we consider only two-dimensional images in this initial study, but the analysis is the same for three-dimensional images.

**Image model**

Let \( Y_n \) denote the time function at pixel \( n \), \( n = 1, 2, \cdots, N \), represented as a vector. We model \( Y_n \) as a random example from a class of time functions associated with some region(s) of the image, each class being defined by a prototype unit vector \( e_k \), \( k = 1, 2, \cdots, K \). Segmentation is achieved by classifying each pixel’s time function. Let \( X_n \) denote the true class label for \( Y_n \). Then the data model is

\[
E \left[ Y_n | X_n = k \right] = \alpha_k e_k ,
\]

where \( \alpha_k \) is the unknown magnitude of \( Y_n \). In addition, assume that the labels \( X_n \) are independent with unknown distribution

\[
P \{ X_n = k \} = p_k , \quad \sum_{k=1}^{K} p_k = 1 .
\]

Due to intraclass variations of the time functions, we use the following similarity metric to quantify the deviation of observed data \( Y_n \) from its associated basis function \( e_{x_n} \):

\[
\rho_n = \frac{Y_n^T e_{x_n}}{|Y_n|} .
\]

Clearly, \( \rho_n \) is affected by the amount of variation in \( Y_n \).

We assume the following truncated exponential distribution for intraclass variation of the time functions:

\[
p(Y_n | X_n = k) = \frac{1}{\gamma_n} \exp \left( -A_n \left( 1 - \rho_n \right) \right) ,
\]

where \( A_n \) is the signal-to-noise ratio (SNR) defined as \( \alpha^2 / \sigma^2 \), and \( \gamma_n \) is a normalizing constant. The probability model in (3) penalizes deviation of the observation vector \( Y_n \) from its class direction, and the degree of penalty varies with the SNR. The intraclass variation in different image pixels is assumed independent.

Our objective is to estimate, given data \( Y_n \), the following underlying parameters of the model: the signal magnitude \( \alpha_n \) and class label \( X_n \) at each pixel, the basis functions \( e_k \), and the probabilities \( p_k \). Compactly, we define \( \Theta = [E, P, A] \) with \( E = [e_1, e_2, \cdots, e_K] \), \( P = [p_1, \cdots, p_K] \), and \( A = [A_1, \cdots, A_K] \).

**EM estimation/segmentation procedure**

We estimate the parameter vector \( \Theta \) according to a maximum likelihood (ML) criterion:

\[
\hat{\Theta} = \arg \max_{\Theta} P \left( Y | \Theta \right) ,
\]

where \( Y \) is the collection of activity functions \( Y_n \).

Given the complex nature of the likelihood function \( P \left( Y | \Theta \right) \), we use an EM algorithm to estimate \( \hat{\Theta} \). For the EM algorithm, the expected log-likelihood function of the complete data is given by

\[
Q(\Theta; \Theta^{-1}) = \sum_{i=1}^{K} \sum_{n=1}^{N} p(X_n = k | Y_n; \Theta^{-1}) \left[ -\log(Y_n) + \left( 1 - \frac{Y_n^T e_{x_n}}{|Y_n|} \right) \log \left( P(X_n = k; \Theta^{-1}) \right) \right] - 2A_n \left( 1 - \frac{Y_n^T e_{x_n}}{|Y_n|} \right) \log \left( P(X_n = k; \Theta^{-1}) \right)
\]

where \( i \) is the iteration index.

The \( E \) (expectation) and \( M \) (maximization) steps of the generalized EM algorithm are then derived from (5). The details are omitted in this summary. The resulting method is termed the soft-clustering algorithm (SCA).
We also derived a simplified version of the EM algorithm by adopting a winner-take-all strategy (assuming \( p_n = \frac{1}{K} \) and \( A_n \to \infty \) in (5)). This algorithm is referred to as wtaSCA.

**Results**

In this study, a single slice (No.70) of the Zubal brain phantom [4] was used to simulate a dynamic study of \([^{11}C]\) carfentanil binding to \(\mu\)-selective opiate receptors. A four-compartment and a three-compartment tracer kinetic model were used to produce time activity curves (TACs) for various brain regions. The model used parameters derived from data in [5] and an input function, blood concentration of tracer, obtained in an actual study. We simulated 23 image frames with a total of four million counts. The pixel size was 4.36mm/pixel and blur had FWHM of 8mm.

The images contain three areas with distinct TACs, shown in Fig.1 (upper-left image). These areas are: (a) background, where no activity is observed (shown in black); (b) areas with almost identical time activity, but differ in amplitude. They correspond to the thalamus, caudate, frontal cortex, temporal cortex, and white matter areas of the brain (shown as gray); and (c) occipital cortex (shown as white).

We compared the proposed approach with the following three well-known segmentation methods: a) Gaussian mixture modeling (GMM) [6]; b) \(k\)-means algorithm [1,7], which is an essentially winner-take-all version of the GMM algorithm; and c) clustered component analysis (CCA) [3].

The classification results for \( K = 3 \) are summarized in Table 1 and Fig.1. In Table 1, the correct classification rate for each class was computed for each area type. From these results, it is clear that the proposed wtaSCA and SCA outperform all the other methods. The wtaSCA is the fastest from a computational point of view. The SCA is slower than the wtaSCA, but faster than the CCA and the GMM algorithms.

**Conclusion**

In this paper we propose a new method for segmentation of dynamic images based on temporal activity functions. The results presented here are very encouraging. In the future we will include the minimum description length (MDL) criterion in the proposed EM framework, which will automatically estimate the model order \( K \). The proposed approach can be used as part of a 4D reconstruction algorithm in which the processing is adapted to the local temporal behavior.

**References:**


