Multi-Feature Landmark-Free Active Appearance Models: Application to Prostate MRI Segmentation

Robert Toth, Anant Madabhushi

Abstract

Active Shape Models (ASM’s) and Active Appearance Models (AAM’s) are popular approaches for medical image segmentation that use shape information to drive the segmentation process. Both approaches rely on image derived landmarks (specified either manually or automatically) to define the object’s shape, which require accurate triangulation and alignment. An alternative approach to modeling shape is the levelset representation, defined as a set of signed distances to the object’s surface. In addition, using multiple image derived attributes (IDA’s) such as gradient information has previously shown to offer improved segmentation results when applied to ASM’s, yet little work has been done exploring IDA’s in the context of AAM’s. In this work, we present a novel AAM methodology that utilizes the levelset implementation to overcome the issues relating to specifying landmarks, and locates the object of interest in a new image using a registration based scheme. Additionally, the framework allows for incorporation of multiple IDA’s. Our multi-feature landmark-free AAM (MFLAAM) utilizes an efficient, intuitive, and accurate algorithm for identifying those IDA’s that will offer the most accurate segmentations. In this work, we evaluate our MFLAAM scheme for the problem of prostate segmentation from T2-w MRI volumes. On a cohort of 108 studies, the levelset MFLAAM yielded a mean Dice accuracy of 88% ± 5%, and a mean surface error of 1.5 mm ±.8 mm with a segmentation time of 150 seconds per volume. In comparison, a state of the art AAM yielded mean Dice and surface error values of 86%±9% and 1.6 mm ±1.0 mm respectively. The differences with respect to our levelset-based MFLAAM model are statistically significant ($p < .05$). In addition, our results were in most cases superior to several recent state of the art prostate MRI segmentation methods.

Index Terms

Active Shape Models, Active Appearance Models, Prostate Segmentation, PCA, Levelsets.

R. Toth and A. Madabhushi are with Rutgers, The State University of New Jersey, Department of Biomedical Engineering.
Multi-Feature Landmark-Free Active Appearance Models: Application to Prostate MRI Segmentation

I. BACKGROUND

In medical imagery, shape model based segmentation has been used in a number of applications including surgical intervention [1], detecting disease within an organ for targeted therapy [2], and for volume estimation [3]. This paper builds on the Active Shape Model (ASM) [4], and Active Appearance Model (AAM) [5] frameworks, two shape based methods commonly used for object segmentation.

A. Overview of ASM’s and AAM’s

The premise of both ASM’s and AAM’s is that a low dimensional representation can accurately describe the shape and intensity appearance of an object. Traditionally, ASM’s define a set of landmarks (specified by their Cartesian coordinates) on the boundary of an object, and Principal Component Analysis (PCA) is performed on the coordinates of the landmarks to yield a statistical shape model (SSM) of the object of interest [4]. Following the generation of the SSM, the intensities surrounding each border landmark are modeled as Gaussian distributions. To segment a new image, boundary locations are automatically ascertained [4], to which the SSM is fit.

However, no information from inside or outside the object of interest is taken into account when using ASM’s. In addition, the shape and appearance have interdependencies, which ASM’s do not consider. To overcome these limitations, the AAM framework was developed [5]. With AAM’s, PCA is first performed on the set of image (2D or 3D) intensities inside the object of interest to generate a low dimensional appearance projection of each training image. A set of low dimensional “linked projections” are then calculated by concatenating the Cartesian coordinates, representing the shape, with the appearance projections, and performing PCA a second time [6]. A linked projection defines both the shape and appearance of an object. To segment a new image, the linked projections are varied, and the original, high dimensional shape and appearance are reconstructed. This process of varying the linked projections is repeated until the reconstructed intensities best match the original intensities [5].

TABLE I: Acronyms commonly used in the paper.

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASM</td>
<td>Active Shape Model</td>
</tr>
<tr>
<td>AAM</td>
<td>Active Appearance Model</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal Component Analysis</td>
</tr>
<tr>
<td>SSM</td>
<td>Statistical Shape Model</td>
</tr>
<tr>
<td>IDA</td>
<td>Image Derived Attribute</td>
</tr>
<tr>
<td>MFLAAM</td>
<td>Multi-Feature Landmark-Free Active Appearance Model</td>
</tr>
<tr>
<td>CLP</td>
<td>Concatenating Low-Dimensional Projections</td>
</tr>
<tr>
<td>CHF</td>
<td>Concatenating High-Dimensional Features</td>
</tr>
</tbody>
</table>

B. Limitations of Existing ASM’s and AAM’s

As stated previously, ASM’s suffer from a number of limitations, some of which were addressed by the development of AAM’s. Several ASM limitations are listed below.

1) ASM’s exclude information regarding the object’s appearance everywhere except the object boundary.
2) ASM’s assume independence of the shape and appearance models.
3) The traditional ASM appearance model assumes a Gaussian distribution for the underlying intensities, an assumption that may not always be valid [7].

In addition, there are several limitations common to both ASM’s and AAM’s. Some of these are listed below.

1) Performing PCA on a set of landmarks may not always accurately capture shape variations in the organ.
2) A large number of anatomical landmarks may be required (usually manually selected) to accurately capture shape variations.
3) Landmark-based models require accurate alignment of an equal number of corresponding landmarks on all training images [4]. To accurately capture the underlying image intensity statistics, each landmark should represent the same anatomical location in each training image [8], [9]. Generating accurate correspondences quickly becomes infeasible on account of the large number of landmarks, and an automated method for landmark detection and alignment can be prone to errors [10].
4) Landmarks require triangulation, and the triangulation algorithm could have a significant computational overhead, and may be prone to errors [11].
5) Both ASM’s and AAM’s traditionally use image intensities [4], [5]. Image derived attributes (IDA’s) such as edge gradients have been previously shown to yield more accurate segmentation results [12].

In this work we extend the traditional AAM framework in two ways. Our multi-feature landmark-free AAM model, denoted as MFLAAM, uses a levelset to capture the shape information [13]. In addition, while the AAM was developed using image intensities, our MFLAAM allows for incorporation of multiple IDA’s (including grayscale intensities), providing the model with additional discriminability.

II. PREVIOUS RELATED WORK AND NOVEL CONTRIBUTIONS

In this section, we provide a brief review of recent attempts to improve ASM’s and AAM’s.
A. Improvements to Statistical Shape Models

A simple linear model may not necessarily be sufficient to accurately capture variations in the object’s shape, and to overcome this limitation a bilinear model could be used to create a SSM [14]. Owing to the previously mentioned problems with the use of landmarks to define SSM’s, some researchers have investigated levelset based representations of object shape, initially proposed by Leventon et al. [13]. This approach involves first performing PCA on a set of minimum signed distances from each pixel to the object’s surface, to yield a set of shape projections. The original levelsets can be reconstructed from these shape projections. It was noted in [13] that reconstructing the levelset from the shape projections will not necessarily result in a signed distance levelset, yet the reconstructed levelsets will be smooth and accurate enough for shape modeling.

This work was later incorporated into the ASM framework by Tsai et al. in 2004 [15] in which the levelset representations of multiple objects were concatenated prior to performing PCA. This allowed for the creation of a set of “linked” projections. In this context, a linked projection is a low dimensional embedding which defines the linear relationship between levelsets of multiple objects. Tsai et al. [15] used this low dimensional linked projection to segment multiple objects simultaneously. However, despite the merits of using levelset-based SSM’s, traditional landmark based SSM’s are still more common [9]. Levelsets were considered as a shape model in the context of AAM’s in [16], in which an AAM was trained to segment out the lateral ventricles on 20 3D MRI volumes.

B. Improvements to Statistical Appearance Models

Complementing image intensities with IDA’s may yield more accurate segmentations [3]. Seghers et al. [17] convolved the intensities of lung CT images with 25 different kernels, and the average Mahalanobis distance over all 25 IDA’s was shown to yield accurate localization of the border. Van Ginneken et al. [18] calculated a Taylor series approximation of image intensities and the optimal IDA’s were then selected and used in an ASM framework.

M. de Bruijne [7] and B. van Ginneken [18] showed that a non-linear k-nearest-neighbor (kNN) based appearance model yields improved segmentation accuracy in terms of both image intensities [7] and IDA’s [18] instead of invoking the normal distribution for modeling the object boundary in traditional ASM’s. These methods demonstrated the utility of using non-Gaussian descriptions of appearance, as well as the usefulness of IDA’s, over simple intensity based appearance models.

In [19], Larsen et al. extended the traditional AAM framework to use wavelet features instead of intensities for 2D images. The image intensities were converted into Haar wavelet projections, mainly as a means to reduce computational cost. The wavelet projections were then used to reconstruct the original image intensities. Ghose et al. [20] employed intelligent wavelet coefficient selection by discarding several wavelet coefficients, to yield an even more efficient, although still highly accurate, implementation of a wavelet based AAM. In addition, Hu et al. [16] used different MR imaging modalities (specifically T1-weighting, T2-weighting, and proton density) as 3 features used to drive an AAM. Finally, Baka et al. [21] showed an AAM derivation which used multiple IDA’s to segment 19 2D cardiac MR images. For more examples of ASM and AAM improvements which are not directly related to the work in this paper, we refer the interested reader to [9].

C. Brief Overview of MFLAAM

The MFLAAM is first trained on a series of images. To segment a new image, the IDA’s are first extracted, and the goal is to calculate the shape. Given a set of linked projections, one can reconstruct an approximation of the original high dimensional data. The MFLAAM is rotated, translated, and scaled, and the IDA’s are reconstructed. The location of the best IDA reconstructions is found, and the shape is reconstructed at this location. The hypothesis is that an accurate IDA reconstruction will correspond to a proper shape reconstruction, and thus a correct segmentation. Figure 1 shows the algorithm for segmenting a new image.

D. Comparison with Closest Related Works

In this work we present the MFLAAM, a framework for incorporating multiple IDA’s into an AAM, and using a levelset instead of landmarks to represent the shape. PCA is used to link multiple objects by first concatenating the objects, and performing PCA on this concatenated space [5], [15]. This is achieved either by (1) Concatenating High-Dimensional Features (CHF), in which one aggregates all the original high dimensional data, similar to Tsai et al. [15], or (2) Concatenating Low-Dimensional Projections (CLP), in which one first performs PCA to reduce the dimensionality of the shape and appearance, and those projections are concatenated, similar to traditional AAM’s [5]. Overall, this work differs from Tsai et al. [15] in that (1) PCA is used to link a levelset and multiple IDA’s instead of multiple levelsets, and (2) PCA is performed using the CLP instead of CHF method.

With each application of PCA, the data is embedded in a lower dimensional space, thereby decreasing the data variance. While the CHF approach involves a single application of PCA, and the CLP approach involves two successive applications of PCA, one would assume that CHF would be the appropriate strategy since it involves lower loss in variance. However, CHF results in a space with extremely high dimensionality, and performing PCA on this high dimensional space can be computationally infeasible. CLP is a computationally tractable method, one employed by the traditional AAM [5], and hence the one we adopt for the MFLAAM. The benefits of performing PCA twice include (1) the ability to process each feature in parallel, and (2) the reduced memory requirements of the Eigen-analysis and covariance matrix calculation. In addition, in Appendix 1 we show that there is only a marginal loss of variance from performing PCA twice.

In addition, while [16] used a levelset in an AAM framework, there are several important differences with the work presented in this paper. (1) The low dimensional projections for the levelsets were constrained to between ±2 standard deviations from the mean shape. While this may be a reasonable
Fig. 1: The MFLAAM segmentation algorithm. The IDA's are transformed and used to calculated the low dimensional linked projections. The linked projections are used to reconstruct an estimate of the transformed high dimensional IDA's. The correlation between the IDA's and reconstructions is maximized. Finally, the linked projections are used to estimate a high dimensional levelset reconstruction and associated segmentation.

assumption for most scenarios, it is entirely possible that a new image might be better segmented with levelsets, where the extent of shape variation falls outside this range. (2) No alignment was performed for the training images in [16]. This is not typically an issue with brain MRI, but critical to address with prostate MRI data, where deformation and the presence or absence of the endorectal coil can cause differences in relative orientations of the gland from patient to patient. (3) No alignment of the AAM model to a new image was performed. Since there is a large region in which the object can appear in a new image, we find that this affine alignment is a crucial step in the MFLAAM algorithm. Without any rotation, translation, or scaling, one assumes that the training levelsets include sufficient pose information, which is not guaranteed. (4) Intensity differences were used to drive the segmentation in [16], whereas we employed normalized cross correlation on account of non-standard intensity values between patient studies in the present work.

We note that IDA's have been previously used in conjunction within an AAM framework [16], [19], [20], [21]. In [19] and [20] the AAM's were developed in conjunction with wavelets, while in [16] and [21] the AAM's were generalized to work in conjunction with any type of IDA's. While [16] and [21] attempted to maximize the similarity between the reconstructions and original features using the $L_2$ norm, whereas the MFLAAM maximizes normalized cross correlation, a useful similarity measure to help overcome intensity non-standardness and extreme intensity values. Another fundamental difference between the MFLAAM and related works [16], [19], [20], [21] is that it employs an explicit feature selection scheme specifically tailored to identifying those attributes that will yield the most accurate shape reconstruction (and therefore most accurate segmentation). Additionally, unlike [21] that only utilized pose information to drive the AAM, the MLAAM utilizes a full range of affine transformations to determine the optimal reconstruction.

E. Application to T2-weighted Prostate MRI Segmentation

The MFLAAM was applied to the difficult task of prostate MRI segmentation. Such segmentation is useful in a computer aided cancer diagnosis system [2], and treatment evaluation via planimetry-based volume estimation [3]. Several segmentation schemes for MR imagery of the prostate have been recently presented, including Klein et al. [22], Martin et al. [23], Pasquier et al. [24], and Makni et al. [25]. Klein et al. [22] performed a registration between an MR image of the prostate and an atlas of training data to achieve a segmentation of the prostate. Martin et al. [23] also used an atlas of training images, but constrained the segmentation model through the use of a statistical shape model. Pasquier et al. [24] used an Active Shape Model [4] method for extracting a statistical shape model of the prostate, which then looked for strong gradients to identify the prostate edge. Finally, Makni et al. [25] used a statistical shape model of the prostate, and clustered the intensities within a manually placed region of interest into 3 clusters: surrounding tissues and fat, central prostate zone, and the peripheral prostate zone. Any pixels within the latter 2 zones were determined to be in the prostate.

In this work we evaluate our MFLAAM scheme in the context of prostate segmentation on T2-w MRI on 108 patient studies. Specifically we compare our scheme against a number of state of the art prostate MRI segmentation schemes in terms of (a) accuracy, (b) computational complexity, and (c) user intervention.

The rest of the paper is organized as follows. Section III describes in detail our training and segmentation methodology. The data and experiments performed are described in Section IV, the results of which are presented in V. We offer concluding remarks and future directions in Section VI.
III. MFLAAM SEGMENTATION ALGORITHM

A. Notation

An image scene is defined as $C = (C, f)$. $C \in \mathbb{R}^P$ represents a set of $P$ pixels, and each $c_k \in C, k \in \{1, \ldots, P\}$ is defined by its Cartesian coordinates $(x, y, z)$. $f(k)$ represents the intensity at pixel $k$. $F_{i,j} \in \mathbb{R}^P$ represents feature $j$ of image $i$, where $F_{i,j} = \{f_{i,j}(k) | k \in \{1, \ldots, P\}\}$ and $f_{i,j}(k)$ represents the value of feature $j$ at pixel $k$ of image $i$. $f_{i,j}(k)$ represents the levelset value at pixel $k$ and $f_{i,j}(k)$ through $f_{i,M}(k)$ represent the $M-1$ IDAs at pixel $k$.

For a segmented image $C_i$, $C^{(1n)} \subset C$ represents an unordered set of pixels inside the object, and $C^{(On)} \subset C$ represents an unordered set of pixels on the surface. Each $c \in C^{(1n)}$ is therefore a pixel inside the object, and each $c \in C^{(On)}$ is a pixel on the surface. A summary of the notation used throughout the paper is illustrated in Table II.

B. MFLAAM Training

The MFLAAM is trained with $N$ images $\{C_1, \ldots, C_N\}$. First, the shape $F_{i,1}$ and the IDA's $F_{i,j}, j > 1$ are computed, followed by the calculation of the feature projections $F_{i,j}$ and linked projections $\tilde{F}_i$.

1) Calculating Shape: The shape $F_{i,1} = \{f_{i,1}(1), \ldots, f_{i,1}(P)\}$ for $C_i$ is represented by the signed distances to the object’s surface [13], and is calculated as,

\[
\tilde{f}_{i,1}(k) = \begin{cases} 
- \min_{c \in C^{(On)}} \|c_k - c\|_2 & \text{if } c_k \in C^{(1n)} \\
+ \min_{c \in C^{(On)}} \|c_k - c\|_2 & \text{if } c_k \notin C^{(1n)}.
\end{cases}
\]

2) Calculating Projections using PCA: Performing PCA on $\{F_{i,1}, \ldots, F_{i,N}\}$ results in $P$ ordered eigenvectors $\psi_j^{(1)}, \ldots, \psi_j^{(P)}$ and associated eigenvalues $\overline{\lambda}_j^{(1)}, \ldots, \overline{\lambda}_j^{(P)}$, where $\overline{\lambda}_j^{(1)} > \ldots > \overline{\lambda}_j^{(P)}$.

Each eigenvector $\psi_j^{(k)}, \forall k \in \{1, \ldots, P\}$ is defined by its $P$ elements

\[
\psi_j^{(k)} = \left\{ \psi_j^{(k)(1)}, \ldots, \psi_j^{(k)(P)} \right\}.
\]

$\tilde{P}$ is chosen to be as small as possible such that

\[
\sum_{k=1}^{\tilde{P}} \overline{\lambda}_j^{(k)} \geq \alpha \cdot \sum_{k=1}^{P} \overline{\lambda}_j^{(k)},
\]

and $\alpha$ is predetermined and $0 < \alpha \leq 1$. The feature projection $\tilde{F}_{i,j} = \{\tilde{f}_{i,j}(1), \ldots, \tilde{f}_{i,j}(\tilde{P})\}$ for image $i$ and feature $j$ is calculated as,

\[
\tilde{f}_{i,j}(k) = \sum_{k=1}^{\tilde{P}} (f_{i,j}(k) - \overline{f}_j(k)) \cdot \psi_j^{(k)}(k),
\]

where $\overline{f}_j(k) = \frac{1}{N} \sum_{i=0}^{N} f_{i,j}(k)$. The feature projections are concatenated as,

\[
\tilde{F}_i = \left\{ \tilde{f}_{i,1}(1), \ldots, \tilde{f}_{i,1}(\tilde{P}_1), \ldots, \tilde{f}_{i,M}(1), \ldots, \tilde{f}_{i,M}(\tilde{P}_M) \right\}.
\]
Each eigenvector \( \tilde{\psi}^{(k)} \), \( \forall k \in \{1, \ldots, \hat{P} \} \) is defined by its \( q \) elements
\[
\tilde{\psi}^{(k)} = \left\{ \tilde{\psi}^{(k)}(1), \ldots, \tilde{\psi}^{(k)}(q) \right\},
\]
which can be rewritten as,
\[
\tilde{\psi}^{(k)} = \left\{ \tilde{\psi}^{(k)}_1(1), \ldots, \tilde{\psi}^{(k)}_1(\hat{P}_1), \ldots, \tilde{\psi}^{(k)}_M(1), \ldots, \tilde{\psi}^{(k)}_M(\hat{P}_M) \right\},
\]
\( \hat{P} \) is chosen to be as small as possible such that,
\[
\sum_{k=1}^{\hat{P}} \lambda^{(k)} \geq \alpha \sum_{k=1}^{q} \lambda^{(k)}. \tag{8}
\]
The linked projections \( \tilde{F}_i = \{ \tilde{f}_i(1), \ldots, \tilde{f}_i(\hat{r}) \} \) are calculated as,
\[
\tilde{f}_i(k) = \sum_{j=1}^{M} \sum_{k=1}^{\hat{P}_j} \tilde{f}_{i,j} \left( \hat{k} \right) \cdot \tilde{\psi}^{(k)}_j \left( \hat{k} \right). \tag{9}
\]
Since the feature projections retain a certain percentage (\( \alpha \)) of the variance, and the linked projections also retain \( \alpha \) of the variance, the total variance retained in the final model is \( \alpha^2 \) of the original variance. This is shown analytically in the Appendix.

**C. MFLAAM Segmentation**

A new, unsegmented image is denoted as \( C_0 \), where \( \hat{F}_{0,j} \), for \( j > 1 \) represents the IDA’s of the new image, and \( \hat{F}_{0,1} \) represents the unknown shape of the new image. The goal is to determine the final segmentation \( C_0^{(T)} \) given the IDA’s. The IDA’s are used to estimate \( \hat{F}_0 \), which can then be used to reconstruct an estimate of \( \hat{F}_{0,j} \) (denoted as \( \hat{F}_{0,j} \in \mathbb{R}^P \)) and yield a final segmentation \( C_0^{(T)} \).

1) **Calculating a Feature Reconstruction (\( \hat{F}_{0,j} \) using IDA’s):** The first step is to calculate a set of linked projections \( \hat{F}_0 \) using the IDA’s. Given linked projections \( \hat{F}_0 \), the feature projections \( \hat{F}_{0,j}, \forall j \in \{1, \ldots, M\} \) are estimated as \( \hat{F}_{0,j} = \{ \hat{f}_{0,j}(1), \ldots, \hat{f}_{0,j}(p_j) \} \).
\[
\hat{f}_{0,j} \left( \hat{k} \right) = \sum_{k=1}^{\hat{P}_j} \tilde{\psi}^{(k)}_j \left( \hat{k} \right) \cdot \hat{f}_0 \left( \hat{k} \right). \tag{10}
\]
The linked projections \( \hat{F}_0 = \{ \hat{f}_0(1), \ldots, \hat{f}_0(\hat{r}) \} \) which minimize the sum of squared differences between \( \hat{F}_{0,j} \) and \( \hat{F}_{0,j} \) are calculated,
\[
\hat{F}_0 = \arg\min_{\hat{F}} \sum_{j=2}^{M} \sum_{k=1}^{\hat{P}_j} \left( \hat{f}_{0,j} \left( \hat{k} \right) - \hat{f}_{0,j} \left( \hat{k} \right) \right)^2. \tag{11}
\]
Equation (9) cannot be used to estimate \( \hat{F}_0 \) directly since we only have the IDA’s (\( j \geq 2 \)) and not the shape (\( j = 1 \)).

The next step is to use \( \hat{F}_0 \) to reconstruct a full \( P \) dimensional feature. Given a set of estimated feature projections \( \hat{F}_{0,j} \), the reconstructed features \( \hat{F}_{0,j} = \{ \hat{f}_{0,j}(1), \ldots, \hat{f}_{0,j}(P_j) \}, j \in \{1, \ldots, M\} \) are calculated as,
\[
\hat{f}_{0,j} \left( \hat{k} \right) = \hat{f}_0 + \sum_{k=1}^{P_j} \tilde{\psi}^{(k)}_j \left( \hat{k} \right) \cdot \hat{f}_{0,j} \left( \hat{k} \right). \tag{12}
\]

The entire process of reconstructing feature \( j \) from the IDA’s is illustrated in Algorithm 1.

**Algorithm 1** ReconstructNewFeature

**Input:** IDA’s \( \hat{F}_{0,2} \) through \( \hat{F}_{0,M} \), feature index \( j \in \{1, \ldots, M\} \)

**Output:** Reconstruction \( \hat{F}_{0,j} \in \mathbb{R}^P \)

1: Calculate \( \hat{F}_{0,j}, \forall j \geq 2 \) using Equation (4);
2: Calculate \( \hat{F}_0 \) using Equation (11);
3: Calculate \( \hat{F}_{0,j} \) using Equation (10);
4: Calculate \( \hat{F}_{0,j} \) using Equation (12);
5: return \( \hat{F}_{0,j} \)

Step 1 in Algorithm 1 calculates the IDA feature projections \( \hat{F}_{0,j} \). Step 2 uses the feature projections \( \hat{F}_{0,j} \) to estimate the linked projections \( \hat{F}_0 \). Step 3 calculates a feature projection estimate \( \hat{F}_{0,j} \) for feature \( j \) using the linked projections \( \hat{F}_0 \). Step 4 calculates the feature reconstruction \( \hat{F}_{0,j} \) using the feature projection estimate \( \hat{F}_{0,j} \). An example of a reconstruction overlaid with the original image is shown in Figure 2.

2) **Segmenting \( C_0 \):** To segment \( C_0 \), a template matching algorithm is employed. The location with the best IDA reconstructions (based on normalized cross correlation (NCC)) are found. \( \hat{F}_{0,1} \) is then reconstructed, and a segmentation \( C_0^{(T)} \) is calculated.

If \( \hat{F}_{0,j}^{(T)} \) represents applying an affine transformation \( T \) to feature \( \hat{F}_{0,j} \), then \( \hat{F}_{0,j}^{(T)} \) is defined as
\[
\hat{F}_{0,j}^{(T)} = \text{ReconstructNewFeature} \left( \hat{F}_{0,2}^{(T)}, \ldots, \hat{F}_{0,M}^{(T)} \right). \tag{13}
\]

After each IDA is transformed \( (F_{0,j}^{(T)}), \) there is an associated reconstructed IDA \( (\hat{F}_{0,j}^{(T)}) \). NCC was used to determine how well each feature was reconstructed, denoted as \( \text{NCC} \left( \hat{F}_{0,j}^{(T)}, F_{0,j}^{(T)} \right) \). A value of 1.0 would represent a transformation that yielded a perfect reconstruction, while lower values represent transformed IDA’s which the MFAAM could not reconstruct from the training data. NCC was chosen as the similarity measure instead of the \( L_2 \) norm because of its ability to overcome intensity non-standardness and its robustness to extreme intensity values. For example, a few extremely bright or extremely dark pixels would contribute to driving the metric if the \( L_2 \) norm was used, while the NCC would not suffer from this limitation due to the fact that it is inherently normalized. The hypothesis is that a high NCC value would occur if the feature is transformed such that the object of interest is in perfect alignment with the trained MFAAM (Figure 3). However, since each IDA is reconstructed
Fig. 2: (a) and (b) represent $F^{(T)}_{θ,j}$ and $\dot{F}^{(T)}_{θ,j}$ for two different patient studies. In both (a) and (b), the original feature image $F^{(T)}_{θ,j}$ is shown as the background image. The reconstruction $\dot{F}^{(T)}_{θ,j}$ resulting from the MFAAM is shown inside the blue box.

Therefore, the average NCC over all IDA’s is maximized to determine the best transformation $T$,

$$T = \arg\max_{T} \frac{1}{M-1} \sum_{j=2}^{M} \text{NCC}\left(\dot{F}^{(T)}_{θ,j}, F^{(T)}_{θ,j}\right)$$  \hspace{1cm} (14)$$

where NCC $\left(\dot{F}^{(T)}_{θ,j}, F^{(T)}_{θ,j}\right)$ represents the normalized cross correlation [26] between $\dot{F}^{(T)}_{θ,j}$ and $F^{(T)}_{θ,j}$. A simplex-based optimization algorithm is used to calculate $T$, which can now be used to calculate the reconstructed levelset $\dot{F}^{(T)}_{θ,1} = \left\{ \dot{F}^{(T)}_{θ,1}(1), \ldots, \dot{F}^{(T)}_{θ,1}(P) \right\}$ using Equation (12). The final segmentation $C^{(In)}_{θ}$ is now calculated as all pixels in which the reconstructed levelset is negative,

$$C^{(In)}_{θ} = \left\{ c_k \mid T^{-1}\left(\dot{F}^{(T)}_{θ,1}(k)\right) < 0 \right\},$$  \hspace{1cm} (15)$$
since the levelset is represented by a signed distance function.
TABLE IV: IDA’s employed with the MFLAAM.

<table>
<thead>
<tr>
<th>j</th>
<th>Name</th>
<th>Operation</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Kirsh (0°)</td>
<td>Convolution</td>
</tr>
<tr>
<td>3</td>
<td>Kirsh (90°)</td>
<td>Convolution</td>
</tr>
<tr>
<td>4</td>
<td>Kirsh (180°)</td>
<td>Convolution</td>
</tr>
<tr>
<td>5</td>
<td>Kirsh (270°)</td>
<td>Convolution</td>
</tr>
<tr>
<td>6</td>
<td>Sobel (0°)</td>
<td>Convolution</td>
</tr>
<tr>
<td>7</td>
<td>Sobel (90°)</td>
<td>Convolution</td>
</tr>
<tr>
<td>8</td>
<td>Gaussian</td>
<td>Convolution (Standard Deviation of 0.5)</td>
</tr>
<tr>
<td>9</td>
<td>Gaussian</td>
<td>Convolution (Standard Deviation of 2.5)</td>
</tr>
<tr>
<td>10</td>
<td>Gaussian</td>
<td>Convolution (Standard Deviation of 4.5)</td>
</tr>
<tr>
<td>11</td>
<td>Variance</td>
<td>Neighborhood of (3 mm)$^3$</td>
</tr>
<tr>
<td>12</td>
<td>Variance</td>
<td>Neighborhood of (8 mm)$^3$</td>
</tr>
<tr>
<td>13</td>
<td>Variance</td>
<td>Neighborhood of (15 mm)$^3$</td>
</tr>
<tr>
<td>14</td>
<td>Median</td>
<td>Neighborhood of (3 mm)$^3$</td>
</tr>
<tr>
<td>15</td>
<td>Median</td>
<td>Neighborhood of (8 mm)$^3$</td>
</tr>
<tr>
<td>16</td>
<td>Median</td>
<td>Neighborhood of (15 mm)$^3$</td>
</tr>
</tbody>
</table>

IV. DATA DESCRIPTION AND EXPERIMENTAL DESIGN

A. Data Description

Our data consists of 108 prostate endorectal MR images, acquired using T2-weighting protocol and a 3.0 Tesla coil. A detailed description of the original dataset is shown in Table III. The prostate capsule boundaries were manually segmented in 3D by an expert radiologist using the 3D Slicer software [27], [28], [29], [30]. In addition, to help determine inter-expert variability a second expert was able to segment a subset of the studies. For 17 studies in which a second expert segmented the prostate capsule, the mean Dice Similarity Coefficient value (Equation (18)) between the experts’ segmentations was 0.899894 with a standard deviation of 0.023272. The raw data for each study was originally 256 × 256 × Z, where 20 < Z ≤ 40. So as not to lose information, the images were upsampled and additional slices were interpolated so that each image was 256 × 256 × 40 pixels (therefore $P \approx 2.6 \times 10^6$). Each image was preprocessed to normalize the intensities and remove the bias field [31]. Since the radiologists acquired additional slices past the apex and base, the user manually selected the first and last slices, which were transformed into the appropriate translation and scale in the Z-direction. This served as the only initialization of the model, and hence the MFLAAM segmentation was minimally supervised.

B. Implementation Details

The images were first affinely aligned to a single study to serve as part of the training. The MFLAAM was implemented in a multi-resolution fashion, with 4 resolutions from 50 × 50 × 40 pixels up to the full resolution of 256 × 256 × 40 pixels. For all experiments, $\alpha = 0.95$ was used, similar to [5].

A 5-fold cross validation was performed, in which the data was split into 5 equal partitions. To train, 4/5 of the images were used, and the trained MFLAAM was used to segment the remaining 1/5 of the images. This was repeated 5 times, until each image had been segmented exactly once.

C. IDA Feature Extraction and Selection

For our experiments, we employed several IDA’s. These included first order grey level features (median and variance of neighborhoods surrounding a given pixel) as well as Kirsch [32] and Sobel [33] IDA’s. Table IV contains a complete list of the IDA’s used for the MFLAAM.

The IDA’s are utilized to estimate the linked projection $\tilde{F}_\theta$ using Equation (11). $\tilde{F}_\theta$ is then employed to reconstruct our final segmentation $\hat{F}_{\theta,1}$. Knowing the true shape $F_{\theta,1}$ would have allowed us to use Equation (9) to estimate $F_{\theta}$. However, $\hat{F}_{\theta,1}$ is unknown, and as such Equation (11) was used. We work under the assumption the difference between using Equations (9) and (11) would be insignificant if the feature projections of the IDA’s and shape are highly correlated. This correlation $R$ can be quantified, and employed for identifying the most discriminating features.

The correlation $R$ between a given IDA feature projection $\hat{f}_{i,j} (\hat{k}_j)$ and shape feature projection $\hat{f}_{i,j} (\hat{k}_1)$ is defined as,

$$R (\hat{k}_1, \hat{k}_j) = \frac{\sum_{i=1}^{N} \hat{f}_{i,1} (\hat{k}_1) \cdot \hat{f}_{i,j} (\hat{k}_j)}{(N-1) \cdot \sqrt{\hat{V}_i \cdot \hat{V}_j}},$$

where $\hat{V}_j$ indicates the variance of $F_{i,j}$ and is described analytically in the Appendix. We were able to compute a score $\phi_j$ defining the correlation between the shape and IDA as,

$$\phi_j = \frac{1}{p_1 \cdot p_2} \sum_{\hat{k}_1=1}^{p_1} \sum_{\hat{k}_1=1}^{p_2} R (\hat{k}_1, \hat{k}_j).$$

The $M$ features with the highest scores are then identified and incorporated into the MFLAAM.

D. Measures to Evaluate Segmentation Performance

The first measure to evaluate the segmentation result is the Dice Similarity Coefficient (DSC). DSC is volume-based and measures the overlap between two segmentations (higher is better), given as,

$$DSC (C_1, C_2) = \frac{|C_1^{(1n)} \cap C_2^{(1n)}|}{|C_1^{(1n)}| + |C_2^{(1n)}|}.$$  

The second measure is the Mean Absolute Distance (MAD). MAD measures the average distance between two surfaces (in mm, lower is better), and is calculated as,

$$MAD (C_1, C_2) = \frac{1}{|C_2^{(0n)}|} \sum_{c_0 \in C_2^{(0n)}} \min_{c_1 \in C_1^{(0n)} \cap C_2^{(0n)}} \|c_0 - c_1\|_2.$$  

E. Experiment $E_1$: Evaluation of Efficiency

This experiment aims to measure runtime efficiency. For this experiment, the number of IDA’s used was varied from $2 \leq M \leq 7$, and the average time per segmentation was noted, as was the average accuracy in terms of DSC.
TABLE III: Detailed description of the data used to test the MFLAAM.

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Image Size (pixels) × Field of View (mm) × Resolution (mm)</th>
<th>MRI Acquisition</th>
</tr>
</thead>
<tbody>
<tr>
<td>T2-weighted, 3.0 Tesla</td>
<td>256 × 256 × Z, 20 ≤ Z ≤ 40 × 140 × 40 ≤ Z &lt; 140 × 0.54 × 3.0</td>
<td>Fast Spin Echo, Endorectal Coil</td>
</tr>
</tbody>
</table>

Fig. 4: Results from experiment $E_1$, which aimed to explore model efficiency. The mean accuracies (a) and runtimes (b) for the MFLAAM are displayed for different number of IDA’s. $M = 2$ represents just one IDA, while $M = 7$ represents six IDA’s.

TABLE V: Models employed in experiments $E_2$ and $E_3$.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Model</th>
<th>Model Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_2 \Omega_{\text{High}}$</td>
<td>MFLAAM using IDA’s with high $\phi_j$</td>
<td></td>
</tr>
<tr>
<td>$E_2 \Omega_{\text{Low}}$</td>
<td>MFLAAM using IDA’s with low $\phi_j$</td>
<td></td>
</tr>
<tr>
<td>$E_2 \Omega_{\text{AAM}}$</td>
<td>MFLAAM with only intensities (no IDA’s)</td>
<td></td>
</tr>
<tr>
<td>$E_3 \Omega_{\text{ASM}}$</td>
<td>Multi-Feature ASM [3]</td>
<td></td>
</tr>
<tr>
<td>$E_3 \Omega_{\text{Base}}$</td>
<td>$\Omega_{\text{High}}$ results from prostate base</td>
<td></td>
</tr>
<tr>
<td>$E_3 \Omega_{\text{Mid}}$</td>
<td>$\Omega_{\text{High}}$ results from prostate midgland</td>
<td></td>
</tr>
<tr>
<td>$E_3 \Omega_{\text{Apex}}$</td>
<td>$\Omega_{\text{High}}$ results from prostate apex</td>
<td></td>
</tr>
</tbody>
</table>

F. Experiment $E_2$: Evaluation of IDA Features

In this experiment, we aimed to determine whether $\phi_j$ is a useful measure for selecting IDA’s. The MFLAAM was run using the IDA’s with the highest $\phi_j$ scores ($\Omega_{\text{High}}$) and lowest $\phi_j$ scores ($\Omega_{\text{Low}}$). Finally, the use of no IDA’s was explored, where $M = 2$ and $f_{i,2}(k)$ represents the intensity at pixel $k$. This is analogous to a traditional AAM, which only uses image intensities, and is therefore denoted $\Omega_{\text{AAM}}$. A list of models are shown in Table V.

G. Experiment $E_3$: Comparison of MFLAAM to other Prostate Segmentation Algorithms

This experiment compares the MFLAAM segmentation accuracy in the context of the prostate MRI segmentation schemes by Klein et al. [22], Martin et al. [34], [23], Pasquier et al. [24], and Makni et al. [25].

In addition, we show the MFLAAM accuracy for different regions of the prostate. The base is considered the first third of the prostate, the midgland the next third, and the apex the final third. The accuracy values are reported independently for each region of the prostate.

As a final test, a multi-feature ASM ($\Omega_{\text{ASM}}$) was constructed, as described in our previous work [3]. For the $\Omega_{\text{ASM}}$, a series of 500 landmarks were placed on the prostate surface in each training image after alignment. Then, a multi-variate, $(M-1)$ dimensional Gaussian distribution was constructed for each landmark point, serving as a unique appearance model for each of the 500 landmarks. To segment a new image, the location with the highest probability was determined for each landmark point, and the landmark-based shape model was optimally fit to these detected locations, as described in [4].

V. RESULTS AND DISCUSSION

A. Experiment $E_1$: Evaluation of Efficiency

The results from experiment $E_1$ are displayed in Figure 4. The runtimes range from just under 2 minutes per volume to approximately 8 minutes per volume (Figure 4b). However, no additional improvement in accuracy was noted after the inclusion of the fifth IDA (Figure 4a). This suggests that after the first four IDA’s, minimal correlation exists with the shape information, or that the subsequent IDA’s chosen by the feature selection scheme had a low signal to noise ratio. Overall, we believe that 3 IDA’s offer a reasonable trade-off between
TABLE VI: Quantitative results in terms of mean, median, and standard deviations of accuracy for $\mathcal{E}_2$ and $\mathcal{E}_3$. Accuracy values for the MFLAAM are reported separately for the base ($\Omega_{\text{Base}}$), midgland ($\Omega_{\text{Mid}}$), and apex ($\Omega_{\text{Apex}}$). Comparison with other state of the art prostate MR segmentation systems in terms of the number of volumes used in the study, the efficiency (in seconds per volume), and the level of user interaction required, ordered by year are also listed. The best results for each measure for each experiment are bolded.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Reference</th>
<th>Volumes</th>
<th>DSC</th>
<th>MAD</th>
<th>Seconds</th>
<th>Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathcal{E}_2$</td>
<td>$\Omega_{\text{High}}$</td>
<td>108</td>
<td>.8766</td>
<td>.8848</td>
<td>.0497</td>
<td>1.51 mm</td>
</tr>
<tr>
<td></td>
<td>$\Omega_{\text{Low}}$</td>
<td>108</td>
<td>.8158</td>
<td>.8673</td>
<td>.1663</td>
<td>2.25 mm</td>
</tr>
<tr>
<td></td>
<td>$\Omega_{\text{AAM}}$</td>
<td>108</td>
<td>.8599</td>
<td>.8804</td>
<td>.0903</td>
<td>1.64 mm</td>
</tr>
<tr>
<td></td>
<td>$\Omega_{\text{ASM}}$</td>
<td>108</td>
<td>.5898</td>
<td>.6196</td>
<td>.1750</td>
<td>4.85 mm</td>
</tr>
<tr>
<td></td>
<td>$\Omega_{\text{Base}}$</td>
<td>108</td>
<td>.8808</td>
<td>.8903</td>
<td>.0596</td>
<td>108</td>
</tr>
<tr>
<td></td>
<td>$\Omega_{\text{Mid}}$</td>
<td>108</td>
<td>.9141</td>
<td>.9238</td>
<td>.0420</td>
<td>154</td>
</tr>
<tr>
<td></td>
<td>$\Omega_{\text{Apex}}$</td>
<td>108</td>
<td>.8424</td>
<td>.8672</td>
<td>.0907</td>
<td>1200 [25]</td>
</tr>
<tr>
<td></td>
<td>Klein et al., 2008 [22]</td>
<td>50</td>
<td>.85</td>
<td>.88</td>
<td>900 [25]</td>
<td>none</td>
</tr>
<tr>
<td></td>
<td>Martin et al., 2010 [23]</td>
<td>36</td>
<td>.84</td>
<td>.87</td>
<td>240 [23]</td>
<td>unknown</td>
</tr>
</tbody>
</table>

accuracy and efficiency. Consequently, we chose to use $M = 4$ for experiments $\mathcal{E}_2$ and $\mathcal{E}_3$.

B. Experiment $\mathcal{E}_2$: Evaluation of IDA Features

The first 3 rows of Table VI show the quantitative results for the $\Omega_{\text{High}}$, $\Omega_{\text{Low}}$, and $\Omega_{\text{AAM}}$. We report separate $p$ values from a 1-tailed paired Student’s t-test for the DSC values between $\Omega_{\text{High}}$, $\Omega_{\text{Low}}$, $\Omega_{\text{AAM}}$, and $\Omega_{\text{Mid}}$. A histogram of the DSC values from $\Omega_{\text{High}}$ over 108 studies is shown in Figure 5.

The value of using $\phi_j$ as a feature selection measure for the MFLAAM can clearly be seen, as $\Omega_{\text{High}}$ performed significantly better than $\Omega_{\text{AAM}}$ ($p = .0473$). When comparing $\Omega_{\text{High}}$ to $\Omega_{\text{Low}}$, the results were even more pronounced, with $p = .000171$. Comparing $\Omega_{\text{Low}}$ to $\Omega_{\text{AAM}}$ resulted in $\Omega_{\text{AAM}}$ being significantly better ($p = .0811$). This suggests that the MFLAAM has the potential to perform significantly better than a traditional AAM when the appropriate IDA’s are selected and used in conjunction with the model.

C. Experiment $\mathcal{E}_3$: Comparison of MFLAAM to Prostate Segmentation Algorithms

The bottom 5 rows of Table VI show the results from the $\Omega_{\text{ASM}}$ as well as other prostate MR segmentation algorithms. The number of prostate volume studies tested in [22], [23], [24], [25] range from 12 to 50 studies, with varying degrees of manual intervention, ranging from completely automated to fully interactive initialization of the segmentation. By comparison, our model is being evaluated on 108 studies and requires only very minimal user interaction. It should be noted that since each of the comparative results operated on different datasets, a direct comparison is impossible. This would involve applying a set of algorithms to the same benchmark dataset, utilizing the same ground truth annotations.

The results show that in a quantitative evaluation involving more than twice the number of patient studies used in either of [22], [23], [24], [25], [34], our model yielded a consistent median and average Dice accuracy of 0.88. This is at least as high, if not higher, than any other state of the art prostate segmentation methods. In addition, our mean absolute distance between surfaces was approximately 1.5 mm, compared to [23], where an error of 2.41 mm was reported.

$\Omega_{\text{ASM}}$ performed poorly, possibly due to many false positive locations in the image which had a similar appearance to the prostate boundary, and took a much longer time to run on a full 3D volume compared to the MFLAAM. The results for different regions of the prostate showed that the MFLAAM performed extremely well in the base and midgland, but most of the inaccuracies were localized to the apex.

D. Qualitative Results

Figure 6 shows qualitative results for two T2-weighted prostate MR images. In both cases, $\Omega_{\text{High}}$ performed signif-
Fig. 6: (a)-(d) represents the results for one T2-w prostate MRI while (f)-(i) represents the results for a second T2-w prostate MRI. The ground truth $C_i^{(In)}$ is shown in (a) in (f). The prostate segmentation results obtained via $\Omega_{High}$ are shown in the (b) and (g), the corresponding results from $\Omega_{Low}$ are shown in (c) and (h), and $\Omega_{AAM}$ in (d) and (i). For (b)-(d) and (g)-(i), the T2-weighted MR image is shown in the background, and the segmentation result is shown as a colored surface (heatmap). Hot colors represent large errors while cooler colors represent small errors between the corresponding model and associated ground truth segmentation. For (b)-(d), red represents an error of 5 mm, while in (g)-(i) red represents 3 mm.

![Fig. 6: (a)-(d) represents the results for one T2-w prostate MRI while (f)-(i) represents the results for a second T2-w prostate MRI.](image)

Fig. 7: Isosurface renderings of the prostate capsule segmentations are shown for $C_1$ (a) through $C_5$ (e), in which the ground truth $C_i^{(In)}$ is shown in green and the segmentation $C_i^{(In)}$ from $\Omega_{High}$ is shown in red.

![Fig. 7: Isosurface renderings of the prostate capsule segmentations are shown for $C_1$ (a) through $C_5$ (e),](image)

icantly better than both $\Omega_{Low}$ and $\Omega_{AAM}$. It can be seen in Figures 6b and 6d, that all models had trouble segmenting the apex of the prostate, but $\Omega_{AAM}$ encountered difficulties with the right side of the prostate. The poor results in the apex are also supported by the results from $\mathcal{E}_3$. $\Omega_{Low}$ in Figure 6c completely missed the peripheral zone, resulting in a DSC value of only 0.62. Examples such as this one lend credence to the necessity for accurate feature selection.

In the second row of Figure 6, all models ($\Omega_{High}$, $\Omega_{Low}$, $\Omega_{AAM}$) performed relatively well. All 3 models encountered difficulties near the levator ani muscles on the left side of the prostate, but this was exacerbated in the $\Omega_{AAM}$ (Figure 6i). However, the $\Omega_{High}$ still performed slightly better than $\Omega_{Low}$, especially near the apical region closest to the endorectal coil. Figure 6 represents the types of errors typically seen in the three models tested in Experiment $\mathcal{E}_2$ ($\Omega_{High}$, $\Omega_{Low}$, $\Omega_{AAM}$) and shows the usefulness of accurate feature selection with the MFLAAM.

VI. CONCLUDING REMARKS

In this paper, we have presented a novel methodology for extending the traditional Active Appearance Model (AAM) framework to include multiple IDA’s, as well as a landmark-free framework for generating a statistical shape model. We have shown that the amount of information lost by using principal component analysis on a series of IDA’s, and combining those IDA’s with a shape model, is minimal compared
to a traditional AAM. This is a significant improvement over current state of the art statistical shape models. Our segmentation algorithm employs an advanced feature selection algorithm, and our final MFLAAM yields both accurate and consistent segmentation results, tested over a large cohort of data. In terms of accuracy, level of interaction, efficiency, and consistency over a large number of volumes, the MFLAAM can output most other prostate MRI segmentation algorithms.

The feature selection algorithm was designed to calculate the most accurate linked projection from a set of IDA’s. One limitation of the current work is that an IDA could yield an inaccurate linked projection yet still result in an accurate segmentation. Future work will explore more advanced feature selection algorithms to address this limitation. In addition, when using the levelset to define the shape, there is an inherent limitation in the fact that all pixels in the image are used. Since the levelset is thresholded at 0, the values of the pixels far from the object boundary (in which \( |f_{1,1}(k)| \) is large) do not contribute to the final segmentation. In future work we will explore weighting higher the contributions of pixels near the boundary. In [35], MR spectroscopy data was used in conjunction with a segmentation scheme, and another facet of future work will entail expanding the MFLAAM to utilize information from multiple protocols (such as T1-weighted and diffusion-weighted MR).

ACKNOWLEDGEMENTS

This work was made possible via grants from the Wallace H. Coulter Foundation, the National Cancer Institute (Grant Nos. R01CA136535-01, R01CA140772-01, and R03CA143991-01), and the Cancer Institute of New Jersey.

REFERENCES

[27] [Online]. Available: http://www.slicer.org


The MFLAAM training involves Concatenating Low-dimensional Projections (CLP) prior to performing PCA. An alternative to the CLP scheme is the Combining High-dimensional Features (CHF) scheme [15]. In the CHF scheme, the original features are simply concatenated and embedded into a low dimensional linked projection. This section explores the additional loss of variance from using CLP versus CHF.

A. Definitions

An object $C \in \mathbb{R}^P$ contains $P$ elements. For each element $k \in \{1, \ldots, P\}$, there is an associated scalar value $f(k)$. Each $C_i$ is associated with a set of $M$ features, $F_{i,j} \in \mathbb{R}^P, j \in \{1, \ldots, M\}$.

The variance $V(S)$ of a set $S = \{C_1, \ldots, C_N\}$ is defined as the accumulated variance of each element,

$$V(S) = \sum_{k=1}^{P} \left( \frac{1}{N-1} \sum_{i=1}^{N} (f_i(k) - \bar{f}(k))^2 \right) = \frac{1}{N-1} \sum_{i=1}^{N} \sum_{k=1}^{P} (f_i(k) - \bar{f}(k))^2$$

(20)

where $\bar{f}(k) = \frac{1}{N} \sum_{i=1}^{N} f_i(k)$ represents the mean of element $k$ across all $N$ objects.

**Definition 1.** Given a set $S_j$ of feature $j$ across $N$ images, $S_j = \{F_{i,j,1}, \ldots, F_{i,j,N}\}$, the corresponding feature variance $V_j$ is defined as $V_j = V(S_j), \forall j \in \{1, \ldots, M\}$.

The total feature variance $V_\Sigma$ is defined as the feature variance accumulated over all $M$ features, $V_\Sigma = \sum_{j=1}^{M} V_j$.

A low dimensional embedding of feature $F_{i,j}$ is called a feature projection and denoted by $\hat{F}_{i,j}$ (Figure 8b). A feature projection $\hat{F}_{i,j} \in \mathbb{R}^P$, where $p_j \ll P$ is defined by its scalar values $\hat{f}_{i,j}(\hat{k})$, $\hat{k} \in \{1, \ldots, p_j\}$. Each feature projection $\hat{F}_{i,j}, \forall j \in \{1, \ldots, M\}$, is required to retain at least $(\alpha \cdot V_j)$ variance, where $\alpha$ is a pre-determined value such that $0 < \alpha \leq 1$. Note $\hat{F}_{i,j}$ is computed independently for each $j$.

**Definition 2.** Given a set $\hat{S}_j$ of $N$ feature projections $\hat{S}_j = \{\hat{F}_{1,j}, \ldots, \hat{F}_{N,j}\}$, the projection variance $\hat{V}_j$ is defined as $\hat{V}_j = V(\hat{S}_j), \forall j \in \{1, \ldots, M\}$.

By definition, $\hat{V}_j \geq \alpha \cdot V_j$, which is rewritten as $\hat{V}_j = \bar{\alpha}_j \cdot V_j$, where $\alpha \leq \bar{\alpha}_j \leq 1$. The total feature projection variance $\hat{V}_\Sigma$ is defined as the accumulated projection variances, $\hat{V}_\Sigma = \sum_{j=1}^{M} \hat{V}_j$.

The dimensionality of $\hat{F}_i$ is given by $q = \sum_{j=1}^{M} p_j$, so $\hat{F}_i \in \mathbb{R}^q$ (Figure 8c).

**Definition 4.** Given a set $\tilde{S}$ of $N$ concatenated projections $\tilde{S} = \{\tilde{F}_1, \ldots, \tilde{F}_N\}$, the total concatenated variance $\tilde{V}_\Sigma$ is defined as $\tilde{V}_\Sigma = V(\tilde{S})$.

After concatenation, a second low dimensional embedding is performed (Figure 8d). A linked projection $\tilde{F}_i \in \mathbb{R}^q$, which is required to retain at least $(\alpha \cdot \tilde{V}_\Sigma)$ variance, is denoted as $\tilde{F}_i \in \mathbb{R}^r$ where $r \leq q$.

**Definition 5.** Given $N$ linked projections $\{\tilde{F}_1, \ldots, \tilde{F}_N\}$, the total linked variance $\tilde{V}_\Sigma$, is defined as $\tilde{V}_\Sigma = V(\{\tilde{F}_1, \ldots, \tilde{F}_N\})$.

By definition, $\tilde{V}_\Sigma \geq \alpha \cdot \tilde{V}_\Sigma$, which can be rewritten as $\tilde{V}_\Sigma = \bar{\alpha} \cdot \tilde{V}_\Sigma$, where $\alpha \leq \bar{\alpha} \leq 1$.

B. Propositions

Propositions 1 and 2 below show that the CLP method (using $\tilde{F}$ to calculate $\hat{F}$) will allow for retention of most of the original variance $V_\Sigma$.

**Proposition 1.** Given $N$ concatenated projections $\{\tilde{F}_1, \ldots, \tilde{F}_N\}$, with a total concatenated variance of $\tilde{V}_\Sigma$, and total feature projection variance of $\tilde{V}_\Sigma$, $\tilde{V}_\Sigma = \tilde{V}_\Sigma$.

**Proof:**

$$\tilde{V}_\Sigma = V(\{\tilde{F}_1, \ldots, \tilde{F}_N\}) = \frac{1}{N-1} \sum_{i=1}^{N} \sum_{j=1}^{M} \left( \tilde{f}_{i,j}(\hat{k}) - \tilde{f}_{i,j}(\hat{k}) \right)^2 = \sum_{j=1}^{M} (\tilde{V}_j) = \tilde{V}_\Sigma$$

$\tilde{V}_\Sigma$ can now be rewritten as,

$$\tilde{V}_\Sigma = \bar{\alpha} \cdot \tilde{V}_\Sigma = \bar{\alpha} \cdot \tilde{V}_\Sigma = \bar{\alpha} \cdot \sum_{j=1}^{M} \tilde{V}_j = \bar{\alpha} \cdot \sum_{j=1}^{M} \bar{\alpha}_j \cdot V_j. \quad (22)$$

**Proposition 2.** Given a total linked variance of $\tilde{V}_\Sigma$ and total feature variance of $\tilde{V}_\Sigma$, $\tilde{V}_\Sigma \geq \alpha^2 \cdot \tilde{V}_\Sigma$.

**Proof:** By definition, $\alpha$ is the lower bound of $\bar{\alpha}$ and $\bar{\alpha}_j$. Substituting $\alpha$ into Equation (22) yields the lower bound of $\tilde{V}_\Sigma$: \[ \tilde{V}_\Sigma = \alpha \cdot \tilde{V}_\Sigma = \alpha \cdot \tilde{V}_\Sigma = \alpha \cdot \sum_{j=1}^{M} \tilde{V}_j = \alpha \cdot \sum_{j=1}^{M} \bar{\alpha}_j \cdot V_j. \]
Fig. 8: The CLP process (computing a set of linked projections from features). (a). $M$ features from $N$ objects are input. (b). Compute a series of feature projections $\hat{F}_{i,j}$. (c). Concatenate feature projections as $\hat{\mathbf{F}}_i$. (d). Calculate a set of linked projections $\tilde{F}_i$ from the concatenated projections. At every step, a certain percentage ($\alpha$) of the data’s variance is required to be retained in the projections. The top row represents the steps performed on each object, while the bottom row represents the variance retained in each step.

As stated previously, the CHF method [15] could have been used to compute $\tilde{F}_i$. A concatenated feature $F_i \in \mathbb{R}^{(P \times M)}$ can be defined as,

$$F_i = \{f_{i,1}(1), \ldots, f_{i,1}(P), \ldots, f_{i,M}(1), \ldots, f_{i,M}(P)\}.$$ (23)

It can be shown that $V(\{F_1, \ldots, F_N\}) = V_\Sigma$ in a manner similar to Proposition (1). However, there are several reasons our MFLAAM uses $\tilde{F}_i$ (the CLP method) instead of $F_i$ (the CHF method) for calculating $F_i$:

1) **Minimal Loss of Data Variance.** Using $F_i$ retains at least $\alpha \cdot V_\Sigma$ variance while using $\tilde{F}_i$ retains at least $\alpha^2 \cdot V_\Sigma$ variance. Since $\lim_{\alpha \to 1} (\alpha \cdot V_\Sigma - \alpha^2 \cdot V_\Sigma) = 0$, the additional loss of variance is minimal if $\alpha \approx 1$.

2) **Computational Efficiency.** The CHF method involves projecting a set of $P \cdot M$ objects to a lower dimensional space. By comparison, the CLP method only requires projecting a set of $P$ dimensional objects. In addition, calculating $F_{i,j}$ can be performed in parallel for each $j$ using the CLP method, which isn’t possible using the CHF method.

$$\tilde{V}_\Sigma = \tilde{\alpha} \cdot \sum_{j=1}^{M} \alpha_j \cdot V_j$$

$$\tilde{V}_\Sigma \geq \alpha \cdot \sum_{j=1}^{M} \alpha \cdot V_j$$

$$\tilde{V}_\Sigma \geq \alpha^2 \cdot \sum_{j=1}^{M} V_j$$

$$\tilde{V}_\Sigma \geq \alpha^2 \cdot V_\Sigma$$