A Hybrid Clustering Pipeline for

Mining Local Patterns in 3D Point Cloud

**Abstract:**

**Introduction:** With recent advances in high-resolution 3D scanning technology and extensive use of LIDAR imaging, 3D point cloud data are now available for many applications. 3D data represent true geometry of real-world objects, which is often compromised in conventional texture of 2D imaging and even in Kinect-based depth sensing. Despite richness in data, 3D point clouds are unstructured and unordered with inconsistencies in point counts, which turns machine learning (ML) of such data structure challenging. Existing literature either maps 3D geometric features (e.g., depth, curvature) onto 2D grid texture~\cite{} or creates stacks of 2D image slices from 3D model ~\cite{} to leverage traditional image recognition pipelines. Such arbitrary 3D to 2D data projections for technical convenience may inadvertently discard useful information of true geometry in 3D data~\cite{Frenet}. Therefore, mining effective patterns directly from 3D point clouds, without introducing lossy 3D-2D data projections, is a nontrivial task.

The task of mining patterns in 3D data has been investigated in objects~\cite{} and human face recognition~\cite{} applications. The analysis of the human face in 3D is an active area of research that facilitates several emerging areas of research, including computational behavioral sciences~\cite{JOLT paper, 3DASD}, biometrics and affective computing~\cite{}, and forensics and medicine~\cite{}. A majority of these applications mine holistic patterns or global shape in 3D to differentiate between a ‘car’ and a ‘table’ or between the face of ‘John’ from that of ‘David’. In contrast, a more challenging task is to recognize local deformation patterns in 3D facial data that differentiate a surprised expression from an angry expression. The task of learning local patterns has been previously addressed by extracting and analyzing changes in local geometric and shape features. However, many of these analyses require grid-based representation of the 3D face that is an approximation of the actual 3D geometry. However, the mining of local patterns directly from raw 3D facial point cloud data is not well explored in the literature. In this paper, we propose to learn local patterns directly from raw 3D point cloud by addressing the following three challenges.

First, 3D point clouds are inconsistent in resolution and the number of points vary from one cloud to another. A common strategy is to sample and interpolate 3D point cloud on to a uniform 2D grid that facilitates operations similar to pixel-based processing. Instead, we propose a distance-based clustering method as a means to directly segment the facial point cloud into distinct and meaningful regions. Second, clustering algorithms often yield different labels for the same region of the point cloud that depends on initial enumeration of the seeds. This makes tracking and identifying the same spatial region with the same label across different point clouds challenging. In this paper, labels for different segments are consistently identified across all face samples to facilitate local pattern classification. Local features mined from these segments are compared across faces for classifying seven prototypical facial expression patterns. Several clustering algorithms have been tested to evaluate facial segmentations. Third, to achieve a constant dimensional feature space for point clouds of varying size, we use parameters after learning a mixture of Gaussian distributions using the expectation maximization algorithm. These parametric features are used to identify most discriminative facial segments for different expression patterns.

**Methods:**

We propose a hybrid of clustering algorithms for learning of local geometric patterns from 3D point clouds of the human face.

**Data set:** The proposed pipeline for analyzing 3D point clouds is evaluated using the Binghamton University 3D facial expressions (BU-3DFE) data set~\cite{}. The data set has 3D faces with seven benchmarked facial expressions of 50 individuals. The seven prototypical expressions are: happiness, anger, fear, sadness, disgust, surprise, and neutral. The 3D facial data are acquired using a high-resolution 3D facial imaging system known as 3dMD~\cite{}. Additionally, the data set provides well cropped 3D faces that are free of outlier points. We extract and subsequently analyze 3D point coordinates (x, y, z) of the provided 3D models in WRL (a virtual reality modeling language extension) format. The class labels provided with the data set are used to perform supervised classification of seven facial expressions.

**Segmentation of 3D point clouds:** Image segmentation is an active area of research that studies similarity among group of pixel intensities or texture features. In contrast to 2D imaging, 3D point clouds form a surface topology with varying point resolution and count. Assuming the symmetric shape of the human face, we segment eight distinct facial regions. First, we employ and evaluate different clustering algorithms and distance metrics to group the 3D facial point clouds into eight facial segments. For facial point cloud segmentation, we use compared the performances of K-means clustering (with Euclidean and Mahalanobis distances), Gaussian Mixture Model, Agglomerative clustering, spectral clustering, and Gaussian mixture model. The algorithm that yields most meaningful facial segments is chosen for the subsequent local pattern analysis.

**Alignment of 3D point clouds:** Registration of 3D point clouds is important, especially when raw 3D coordinate data are used for the pattern analysis. We perform the registration at two levels: 1) the nose tip is identified as the highest Z-axis value and this coordinate point is used to align and adjust all faces such that nose tip is the origin of the 3D Cartesian coordinate and 2) after segmentation of the facial patches, all patches are aligned at a common patch centroid.

**Spectral clustering for dimensionality reduction:** The spectral clustering algorithm first obtains the Laplacian matrix L by subtracting the adjacency matrix A from the degree matrix D. Here, A is a symmetric binary matrix with entries 1 when a point pair (i, j) is adjacent and 0 otherwise. The diagonal matrix D contains the degrees (number of edges incident to a particular point from other points) of individual points in the patch. The diagonal entry for (i, i) is obtained by summing all columns of the i-th row in A matrix.

For a 3D point cloud patch with N points, A, D, and L are NxN matrices. The Eigen value decomposition of the L matrix yields Eigen vectors, V of \Re^N. Then, V\_λ = { x ∈ \Re^N : Ax = λx } is a subspace of V for λ ∈ R, where λ is the Eigen value. In spectral clustering, the first m Eigen vectors are chosen corresponding to m lowest Eigen values. Thus, a low-dimensional (m<<N) subspace is obtained for N number of points in the patch. We choose m<=3 for our study.

**Parameter learning by subspace clustering** The m-dimensional spectral features of each surface patch are grouped into Q clusters using Q m-dimensional Gaussian distributions. The Gaussian distribution parameters are learned via the expectation maximization (EM) algorithm. The EM algorithm learns a set of Gaussian parameter vectors: S = {w, mu, sigma} for each cluster distribution that maximizes the likelihood function P(x|ci) = exp {} using a posterior probability estimate P(ci|x) as below.

There will be Q weights (w) for Q Gaussian clusters. The m-dimensional mean vector (\mu) will be estimated for each of the Q Gaussians, which yields a total of mXQ parameters. We extract the m-dimensional diagonal vector from the mxm covariance matrix (\sigma). Total covariance parameters for each patch will be mXQ. Each patch will yield a total of mQ + mQ + Q or Q(2m + 1) parameters. Thus, each 3D face will be represented by 8\*Q\*(2m+1) features.

**Selecting features and facial segments:** The 120-dimensional feature vector will be subjected to feature selection. We identify and discard the features with lowest variances across seven 3D faces with different expressions of the same individual. Features with low or zero variability suggest that those are not sensitive to local deformation in the face. The selected features with highest variances across facial expressions can be used to trace back which segments of the face are most sensitive to facial expressions. Finally, selected features from facial segments will be used to visualize class separation in 2D for seven facial expressions.