Real Time Estimation and Prediction of Similarity in Human Activity Using Factor Oracle Algorithm

by

Sudarshan Prashanth Seshasayee

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved April 2016 by the
Graduate Supervisory Committee

Xin Wei Sha, Co-Chair
Pavan Turaga, Co-Chair
David Tinapple

ARIZONA STATE UNIVERSITY

May 2016

ABSTRACT

The human motion is defined as an amalgamation of several physical traits such as bipedal locomotion, posture and manual dexterity, and mental expectation. In addition to the "positive" body form defined by these traits, casting light on the body produces a "negative" of the body: its shadow. We often interchangeably use with silhouettes in the place of shadow to emphasize indifference to interior features. In a manner of speaking, the shadow is an alter ego that imitates the individual.

The principal value of shadow is its non-invasive behaviour of reflecting precisely the actions of the individual it is attached to. Nonetheless we can still think of the body's shadow not as the body but its alter ego.

Based on this premise, my thesis creates an experiential system that extracts the data related to the contour of your human shape and gives it a texture and life of its own, so as to emulate your movements and postures, and to be your extension. In technical terms, my thesis extracts abstraction from a pre-indexed database that could be generated from an offline data set or in real time to complement these actions of a user in front of a low-cost optical motion capture device like the Microsoft Kinect. This notion could be the system's interpretation of the action which creates modularized art through the abstraction's 'similarity' to the live action.

Through my research, I have developed a stable system that tackles various connotations associated with shadows and the need to determine the ideal features that contribute to the relevance of the actions performed. The implication of Factor Oracle [3] pattern interpretation is tested with a feature bin of videos. The system also is flexible towards several methods of Nearest Neighbours searches and a machine

learning module to derive the same output. The overall purpose is to establish this in real time and provide a constant feedback to the user. This can be expanded to handle larger dynamic data.

In addition to estimating human actions, my thesis best tries to test various Nearest Neighbour search methods in real time depending upon the data stream. This provides a basis to understand varying parameters that complement human activity recognition and feature matching in real time.

This piece of work is dedicated to my loving grandparents,

Chandra & Sampath,

They showed their legacy a way of living.

They created a legacy of loving people I call family.

They have always hoped the best for me and always will.

ACKNOWLEDGMENTS

‘Mata, Pita, Guru, Dhaivam’- In Sanskrit literally translates to ‘Mother, Father, Teacher, God’- The order of being I owe everything to.

Padmini, my mother, without whom I too might have been a shadow. Seshasayee, my father who taught me to be an invisible superhuman.

Dr. Xin Wei Sha, who so graciously took me under his wings and helped me feel a sense of accomplishment. He constantly pushed me to seek a new perspective to research topics. He inspired me to experiment with representative technology. He also encouraged my adventures trying to implement the most radical ideas. He kindled the fire of imagination and taught me how to keep it lit.

Dr. Pavan Turaga, my technical conductor. He compelled me to delve into topics of my interest and schooled me. He taught me the value of making suggestions and backing it up with resolute results. He also inspired me to take on tasks beyond my capability and to master it.

David Tinapple who taught me like a friend. His methods of ensuring an experience through hands-on applications aided my confidence in those topics. He taught me to look at the root of questions and build upon it. He inspired me and gave me a fresh perspective on life by simplifying the most complex of systems without breaking a sweat. He is the ‘Cool Cucumber’ personified.

My sister Janani, who finds the quirkiest ways to inspire me. My dear friend Sharon because of whom my dormant creativity was discovered and was crucial in helping me decide upon my Masters. My aunt Sujatha, who has always the right words to

push me towards my goals. Never lets me back down from a challenge. I owe much of my will-power to her.

I have a huge list of friends who I thank for putting up with my incessant ranting while motivating me to still move forward. Joshua, Skaidra and Jennifer taught me the art of taking breaks in between work. Varsha, with whom I have spent 2 years trying to figure out the meaning of an artsy engineer.

Ira A. Fulton School of Engineering for helping me tap my connection with electronics and providing a platform to leverage the concepts I have learnt every step of the way through innovative applications. Also for the strange concentration affiliated with Arts, Media & Engineering. The sole reason I cultivated a wide range of skills and found my true calling.

I am grateful to School of Arts, Media & Engineering for being a second home and the faculty my extended family. It was here I found like-minded people and a stage to mix my knowledge of technology and aspirations to create art. I also owe much of my nascent thinking about art, research, and technology to the Synthesis Centre and affiliates at the Topological Media Lab.

I am very thankful to Arizona State University for understanding the need for interdisciplinary and promoting that goal.

I am very grateful to the institutions for giving me the opportunity travelling from the other side of the globe. The generosity in part made my degree possible financially in addition to providing a pragmatic outlook on life through academia.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER 1: INTRODUCTION

1.1 Motivation

The concept of shadows being a significant extension of the human anatomy is in principle a metaphorical relation. Optics and light suggest a metaphorical environment around a subject. For example, a given tonality of ambient light can be perceived by some observers as sunrise, by others as sunset. Among the most evocative effects of light on a human body in fact can be, the silhouette cast by the body. This shadow appears ‘void’ with no life of its own, yet it is wedded to the living body since it retains all physical traits associated with the same subject.

Shadow puppet theatre has been used throughout Asia for centuries. They incorporated folklore and myths in unison with shadow puppets to tell tales. Laufer, Berthold in Oriental Theatricals: The history of shadow play [15], explains its evolution through the generations. The Han Dynasty was at the root of this art form’s evolution. This art form later spread to India and other parts of the Asian subcontinent.

My research builds from this strong cultural motivation: in my thesis I have proposed several methods to interpret human motion. At the core of the computation, I test the efficiency of various Nearest Neighbour (NN) search methods on the Factor Oracle Algorithm (FOA).

Building upon this goal to estimate human motion, I have developed a system that accommodates various modes for feature definition, segmentation and classification. To elucidate the same, the system can define features in two ways. First, using skeleton joint point data helps us understand the Inverse Kinematics (IK) of the

subject. The joint point tells us the exact path taken by the action performed by the subject. Second, the optical flow of the subject is calculated. This enables us to utilize the quantity of motion in contrast to IK quality of motion. IK data provides the exact spatial position of movement thus providing information about how pristine the action was performed. While optical flow tells the net motion within a set window of frames.

Consequently, the set of features can be defined as a point in a high dimensional space. Through compression techniques, these features are mapped to lower dimensions while keeping data information loss to minimum. Further, the system also emphasizes the study of techniques in high dimensional space. The feature space also introduces the concept of dense trajectories that contribute to these dimensional feature spaces.

As technological advances are made, one-to-one mapping of joints to characters in stories, adds dynamic control. Skilled puppeteers interpret these shadow dolls as an extension of their body. Thus, allowing a high degree of mapping with IK. V.N. Edward in his work about Javanese Wayang Kulit [28] explains the study of these joints as a significant factor when creating these dolls.

It is necessary to expand our knowledge of motion. The sub-level interpretation of motion can be viewed as the interactivity of various parts of our body with one another. This leads to an intrinsic interpretation of how our body understands activity. This can also be perceived as an extrinsic phenomenon. Where the amount of motion creates clarity in its subjective existence in that localized space.

Further, my research allows me to create flexibility by incorporating algorithmic perception to these mappings. In the process, developing a portrayal of a finite state

automaton that imitates the subject and complements his/her action with a similar one. The FOA is the core construct that allows the computation to estimate this similar action. However, it is possible to substitute the pattern search algorithm to accommodate the incoming data. As new alternates try to adjust with the system, the definition of similarity changes accordingly to the context.

The similarity of the action is extracted from a database within a marginal error. The playback of these shadow interpretations is in linear time and space. The perceptual algorithm picks the nearest bin of feature vector as parsed from its data space and strings together a sequence of frames that result in the playback loop.

Similarity is a fuzzy word. It brings into question how motion can be accommodated for similarity. Can the action of moving your arms up and down signify the growth of a plant? Or to wave Hi be used to signify to wave Bye? What are the parameters that are taken into consideration to assume this nomenclature? This opens up a wide non-exhaustive research into social psychology. There is no definitive conclusion regarding the contextual usage of this gestural behaviour. It varies due to several factors. This ranges from the culture the subject was exposed to, to the subject's command over the language. McNeill in his book Language and Gesture tries to define the line that separates between presenting one's abstract thought using gestures and a pragmatic mode that means or symbolizes a sentence structure.

To test the system for estimating human action using FOA, my thesis branches out into a performance module and an art installation. While the research tests the efficiency of the hybrid algorithms, the installation describes the presentation of the system. The former creates a ground to train a computer and create these structures that define a framework which makes contextual sense. The latter, as an art

installation adds a new dimension of concrete presentation of experience. This creates elements that are open to expressive interpretation.

While I mainly focus on developing a system that encapsulates a method to recognize actions, the artistic component is to provide a platform to look at the experiential potential of the system and not to research into psychological trends through the same.

To understand such linguistic flaws, through further research, I aim to train a network that tries to understand these contexts. My interpretation of similarity through these works is not the structural resemblance of movement but to imply connect between the breadth of the motion and its meaning.

CHAPTER 2: LITERATURE REVIEW

2.1 The Philosophy

Human meaning of a sign whether it's some alphabetical characters, a figure or a moving shadow depends on the context. Ludwig Wittgenstein in 'Philosophical investigations' [31] provides the basis to explain the contextual difference of a word, a phrase or an image.

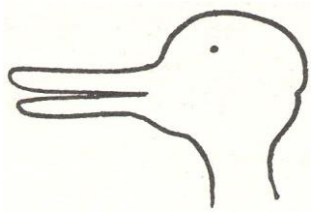

Figure 2.1 Duck-Rabbit, 'Philosophical Investigations'

The figure above as defined by Wittgenstein, can be perceived as a duck, while another would argue this is a rabbit. These challenges provide a primary paradigm to depict how our mind perceives an image. Extrapolating on the same principle, what if every time we looked at a shadow, it looks different; it does different things; it has a motive of its own. Changing the context can alter the primary definition of shadow.

Wittgenstein goes on to break down other concepts pertaining to a sentence structure [32]. He tries to analyze if the sentence makes an associative connotation or does a word in that sentence suffice to bring about the same significance. Drawing upon a similar implication does the repetitive motion symbolize a gesture or is a singular non-repetitive stroke sufficient to indicate the same.

Additionally, he also tries to define the purpose of reference or anchor points in his argument. There needs to be a unanimous structural similarity between the subject and the predicate. To a degree, this helps tackle the fuzziness of the definition of the word ‘similarity’. If a system can self-generate a sequence of fluent interpretations, the subject must derive meaningful inferences from these suggestions. Else, the system fails to create a cognitive experience.

This leads to the study of self-sustaining processes or autopoiesis. Maturana and Varela [9] in their work on autopoiesis and cognition established a reasonable connection between a system architecture and sociology. They defined the functionality of autopoietic machines; to paraphrase, a system that provides a constant feedback to the network and also constitutes it. To understand the full length of its application with shadows, it is sufficient to know that these autopoietic machines have individuality and their operations specify their boundaries. Individuality expands to state that the behaviour of the system is independent of the observing user; however, the interacting user is part of the system at the organization level. Its operation states that if a system is defined by its degree of interaction, the resultant is also proportional to that interaction. To create a context for this interaction, it is essential to implement a module which in addition to estimating the interaction in the system, learns a pattern that poetically follows the user.

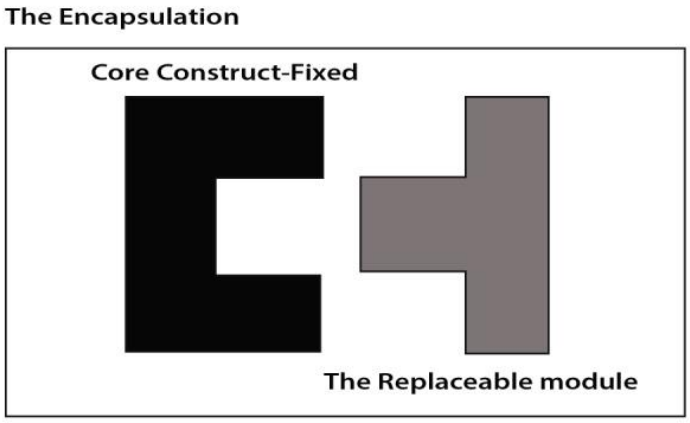

Figure 2.2. The hierarchical chart of the proposed software framework

2.2 The Core Construct

Following up from the philosophical outlook on this system, Allauzen et. al [3] proposed a new structure for pattern matching in their paper. This paper analyzes the factors in an acyclic method and follows a linear transition. This phenomenon is known as the Factor Oracle Algorithm (FOA). Their paper provided an insight into understanding the method of deriving the next state through minimal computation and recognizes the factors of the scrubbing window of feature vectors. They aimed to derive clear models to match strings from previously known lexical data. The concept of Backward Oracle Matching was introduced and tested against various other algorithms such as the Boyer-Moore algorithm. As a fixed length window scrubbed through the linear time line, the m+1 state was concatenated with the previous m states to provide a state of continuity. The failure of finding a string match within the window led to looking forward and probabilistic model with and Expectation governed by a random variable ṗ. This provided a general understanding for coding the algorithm. However, to extrapolate this algorithm and it's feasibility of real-time video processing further research was needed.

This opened up the study of its applications where there is a need for real-time computation and it did not use indexed data to iterate through yet provided constructive feedback. Assayag et. al modelled OMax, a dynamic topology for improvisational learning [4]. This extensively applies FOA in real-time music analysis. They use Max as the feed forward system and OpenMusic as a feedback loop to implement the statistical model. This allows the Max patch to improvise over the performer's playability by reading MiDi values and controlled by an OSC protocol to transmit data to OpenMusic and adapt to the performer.

2.3 The Replaceable

The underlying concept that governs pattern matching using FOA is to find the nearest feature neighbour ahead of it in the list of sequential frames. With high-dimensional vector space, the recursive analysis to fit a vector with a child node needs to be cost-effective. Thus, a trade-off between precision and time is essential. This requires the study of various approaches to Nearest Neighbour (NN) search methods.

R. Kurniawati et. al in their paper [14] provide an efficient NN space partitioning approach using weighted Euclidean metrics which effectively implements a branch and bound methodology. This can improve the complexity from $O(n^2)$ to $O(n)$ depending upon the orientation of the bounding hyper-box since the partition criteria are at the intersection of the two dimensions. This happens only when the weight matrix changes. The approach uses R-trees which not only uses Euclidean distance but other distances as well. This is a useful approach when the testing and training are both known and structured. Additionally, this looks at the incoming frame as a 2D matrix. To extrapolate this approach to *k*-dimension feature vector exponentially

increases the complexity to 2^k. This gives up the time bound constraint. In my proposed system, I aim to explore hyperplanes towards unstructured data.

On further research, Arya, et. al [22] propose a variation to the space partitioning approach. The method suggests the use of a $(1+\varepsilon)$ approximate nearest neighbour. The authors also suggest a priority queue to speed up the search. Such an approach is known as "error bound" approach and soon become less efficient on increasing the dimension k. This is a favourable approach as long as $k < 25$. Outside this boundary condition, it proves to be costly even though it is time efficient. Due to the curse of dimensionality, the proposed system works under 20 dimensions.

An alternate method of incorporating nearest neighbour searches is using hashing. The best approach to do the same is locality sensitive hashing (LSH) [1]. It uses hash functions to determine its proximity of hashes of elements that are likely to be close to each other. However, experimental results show that, due to the cost of segregating the data into hashes and then determining the proximity being high, they are outperformed by space partitioning structures such as a k-d tree.

To understand the variations to a k-d tree, M.Muja and D. Lowe provide a brief understanding of scalable nearest neighbour algorithms [17] such as Randomized k-d tree and priority search k-means tree. They propose a method to randomly search several k-d trees at the same time. While this can only partition the data 1 dimension at a time, an alternate approach suggests that using k-means clustering based on priority considers the entire length of a feature vector and also then uses k-d tree thus significantly reducing search space and load. In this paper, they also explain the

development of their FLANN library which chooses the most efficient NN approach depending upon the structure of the dataset being analyzed.

Self-Organizing Maps (SOM) is another interesting concept that allows to the input sample to be discretized in a competitive learning manner. In the works of W. Huang et. al [29], the use of SOM is to classify human activity. However, it is highly efficient while predicting against identical models of Jia et. al [13] and Ali et. al [22]. This throws light on a new approach that can be adapted for pattern matching using a neural network. This uses Euclidian distance to find the best matching unit (BMU) and updates weights of that unit and its neighbour using a monotonically decreasing learning rate γ.

$$w_{v_{i'}} = w_{v_i} + d\gamma(v_i - w_{v_i})$$

Where LHS is the weight of the next index i' which is updated by the current weight and BMU distance d. v_i is the sample vector. As the weights are updated, a significant separating boundary is established. This threshold creates an array of vectors that incorporate the longest common sequence scheme to determine the trajectories for the entire set of frames. This is definitely “space bound” and is similar to the limitations faced with the randomized k-d tree method.

2.4 Encapsulation

Preceding these NN approaches and drawing analogies into video processing, requires a stable platform to obtain high dimensional data with ease and generate high performance on a complex algorithm. This opened the study of the previous video

based Human Activity Recognition(HAR) analysis [11]. The paper by Shian et. al addresses core technologies and applications from low-level to high-level representation. It also tackles the control system for human activity recognition, namely segmentation, feature extraction and classification. It serves a significant guide to the state of the art approaches towards the goal of HAR. It is a non-exhaustive list of approaches where it is inconsequential to randomly pick one from each category and expect the hybrid model to work. A thorough explanation is also provided about the functionality of each module. Depending on the necessity of the application, one can deduce a combination of these modules.

The author goes on to talk about the various approaches inside

- Segmentation: Optical device, Background subtraction, Gaussian Mixture Model(GMM), Statistical Model, Tracking, Optical flow, Temporal difference etc
- Feature Extraction: Discrete Fourier Transform(DFT), SIFT features, Histogram of Gradients(HOG) features, Appearance-based, body modelling
- Classification: Dynamic Time Warping (DTW), Hidden Markov Models (HMM), Support Vector Machine (SVM), k-Nearest Neighbour (KNN) etc.

We can study these models further and test our application's efficiency against various reasonable combinations. The above-mentioned sub-topics are useful techniques in computer vision and usually adopts a non-linear approach to tackling these features that define an activity.

Activity recognition techniques of today involve the acquisition of enormous amounts of redundant data followed my feature extraction of a rich set of features. This follows high-cost computations to render these features and recognize a set of labelled activities. There is the need for compromise with one of these steps to create a more real-time environment. Thus, providing a faster, closer to real-time functionality to analyze the captured data.

Using a Microsoft Kinect as a low optical device to get an input stream of videos to analyze became increasingly difficult since Apple bought the rights to OpenNI. This led the open source developers to work around on unstable builds of the same. As one tries to work with the depth map on the device, there is a significant noise with the subject's real shadow due to lighting and another natural phenomenon. Gabriel Danciu et. al in their paper about shadow removal in depth images [6] talk about ways to normalize the data even before applying high-level perceptions such as segmentation etc.

However, to understand the removal of real shadows and implement the augmented shadows, it is necessary to decide on the data. This includes the parameters that need to be pre-processed, cleaned and normalized before performing any of these research modules. One such instance is the percentage of shadow [19] needed to be exposed or applied upon the subject. This provided an insightful case study on the progress of computational algorithm within animating libraries. The paper tests its algorithm on 4 test cases and analyzes its efficiency over several epochs and also the strain on the system to compute the same.

As we delve into recreating shadows that closest resemble the user we start to realize that it becomes a heavy load on the system, the algorithm and graphics to create a shadow that convinces the user as their own. The paper on silhouette maps [25] is to derive an approach with structured data. However, the system suffers in performance when unstructured data need to be processed in real-time.

An alternative approach to research on current trends throws light upon the principle of trajectories. This helps to analyze motion patterns. H. Wang et. al in their paper on dense trajectories and motion descriptors [30] explain how various vectors contribute to gauging the overall feature space. Including optical flow analysis on KLT and SIFT filters such as HOG and MBH features. This provides a basis on how to go about and develop a pool to match 'similarity'. However, other techniques need to be adapted to improve the efficiency of this model.

N. Sundaram et. al provides an alternate strategy to complement dense trajectory descriptors by introducing the concept of large displacement optical flow [27]. This variational technique finds discrete points and equates it to continuous energy formulation and helps better analyze on a feature vector that has large movements. However, for smaller movements that could create a pattern match with directionally dissimilar videos, this would fail. Thus, it was reasoned against the use of large displacement.

While [11] & [12] implement it with an SVM classifier, the study calls for a succinct test on my framework with another classifier such as HMM. Eunju Kim et. al provides a perception of these models as applied in pattern discovery towards human

activity recognition [12]. The use of a classifier allows the system to create a branch of study that in addition to estimating human activity using an offline data set, serves as a classifier to segregate activity using an on-line dataset.

Support vector machines (SVM) are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. Aizerman et. al [2] propose a method to create non-linear classifiers. Given a set of training samples, each marked as belonging to one of two categories (0 or 1); an SVM training algorithm builds a model that assigns new samples into one category or the other, making it a non-probabilistic binary linear classifier.

An SVM model is a representation of the samples as points in space, mapped so that the samples of the separate categories are divided by a clear gap that is as wide as possible. New samples are then mapped into that same space and predicted to belong to a category based on which side of the decision boundary they fall on in the domain space. The principle behind this model is to maximize the distance between the binary classes. The largest distance equal from both classes is to be considered while implementing the model.

While there has been extensive work in the field of feature matching, human activity recognition and pattern learning, my thesis is a framework of several combinations of these concepts. This also provides insight into the efficiency of a model in fusion with another. Additionally, the application of FOA has not been extensively explored in video processing and human activity recognition. My thesis proposes to apply FOA

with a new perception on NN searches. Further, my thesis also expounds the system that encapsulates a self-sustaining structure that can be adapted for future branching of this algorithm.

CHAPTER 3: METHODS

3.1 Overview

In my thesis, I present a system that draws upon many of the works referenced and propounded above and propose a revised method to implement FOA outside its intended purpose in text semantics or predictive design in music technology. The system aims to tackle contextual interpretations using video processing and hybrid model for nearest neighbour search with sequence matching.

I have expanded the principle of my thesis over the next three chapters. These chapters diversify the application of each element in the system that analyzes a pattern of human motion and recommends a ‘similar’ motion. These are defined over a number of features and provide ample validation for these features.

The use of C++ and OpenCV suggest the flexibility with which one can experiment with the system. While C++ is a programming language capable of handling large computations, OpenCV is powerful software that was designed to experiment with image and video processing.

There are many toolboxes inside C++ that allow object oriented programming and permit handling a large amount of data. OpenCV has built-in modules for machine learning and stochastic processes that simplify the quantity of effort towards reinventing the wheel for video compression and analysis.

Through the rest of my thesis, the methods applied include the use and efficiency of dense trajectories using the LK filter and extract a useful high dimensional feature vector from this in real time. The system also allows the reference dataset to be offline or created in real-time. Further, a dimensionality reduction algorithm is applied for

pattern matching in the FOA. This allows a smaller space to then suggest a set of 'similar' sequences that complement the human action.

As the computational dependency increases, a perceptual coding approach was induced to modularize the data and reduce memory exploitation.

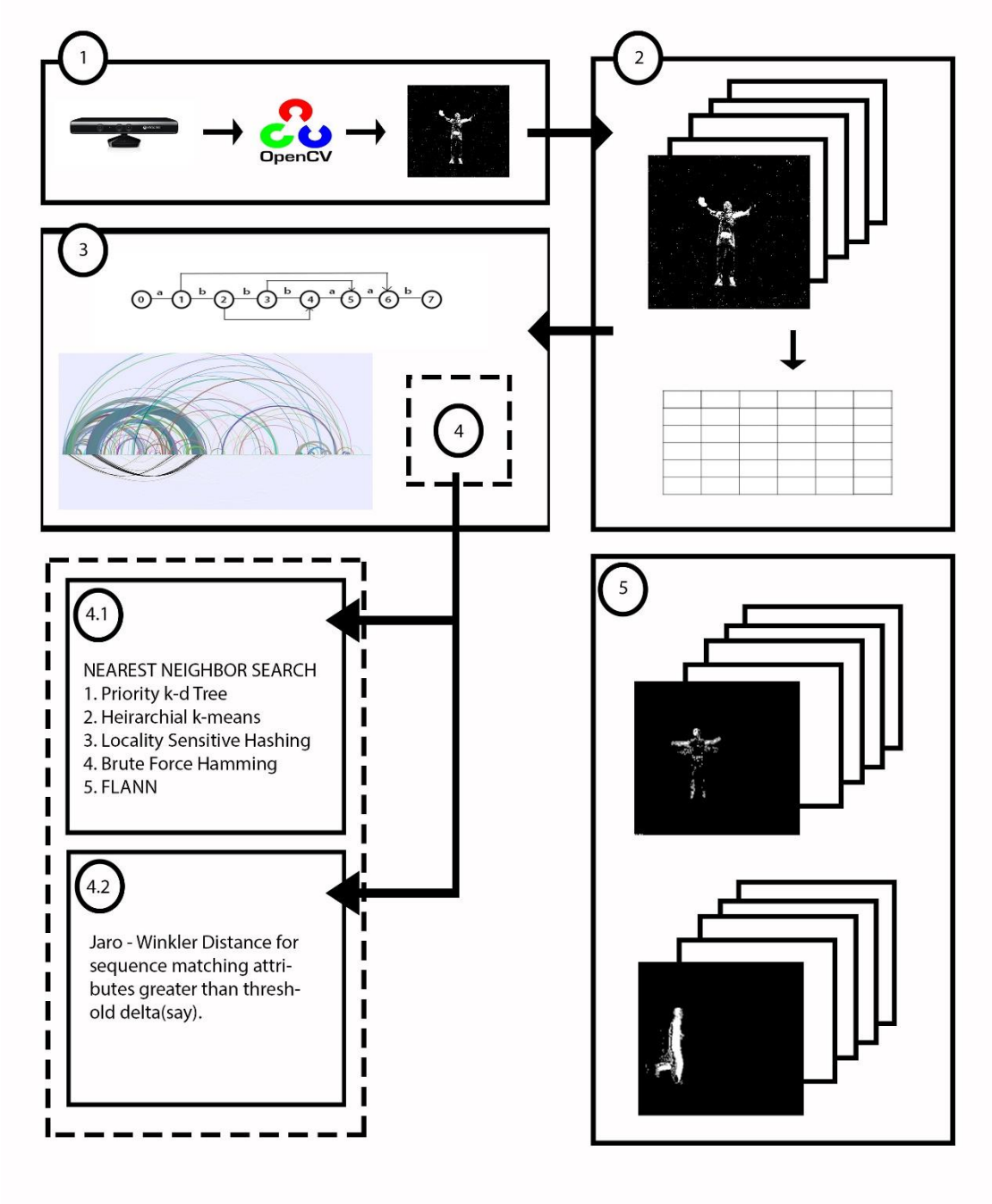

Figure 3.1 An overview of the system architecture to better understand the flow

With reference to the above figure, stage 1 is indexed as the segmentation phase. This utilizes the maturity of OpenCV with a Kinect to obtain a background subtracted model using expectation maximization algorithm in Gaussian mixture models. This is super imposed with optical flow feed from the system and populates the database with feature vectors in stage 2. This introduces the core construct of my system, the FOA which is trivially defined as a word graph and ostentatiously by the visualization such as in stage 3 derived from womax [4] in the above image. This system encompasses the technical disposition to use nearest neighbour search algorithm which could be replaced and modified with one or another approach. This behaves like a shared object. As the module can accommodate any of the conventional approach, I propose a hybrid of two such approaches. This overlooks the drawbacks of FLANN libraries to only pick a 1-1 method to optimize the search space. Stage 4 thus has 4.1 and 4.2 to explain that the proposed system not only fuses two NNS phenomenon but incorporates a sequence match algorithm known as Jaro-Winkler distance that further reduces the noise through the error bound approach. Lastly in stage 5, the feature vector and its derived pattern has a 1st order retrieval which closes matches the pattern and based on priority retrieves the 2nd order or predicted patterns that satisfy the criteria for NNS algorithm.

3.2 Segmentation

3.2.1 Static Camera

We use a standard Kinect camera model 1414. With an IR emitter and IR sensor, we are able to obtain the depth and colour image of the subject in the scene. Further, we place the camera in a stable location. This way the background, light and other

affecting parameters are previously accounted for. This helps to differentiate the foreground subject from the background and makes segmentation easier.

3.2.2 Background subtraction

With a stationary background, it becomes clear that any motion detected is due to the moving subject in the foreground. We extract the depth map from the Kinect and segment the subject in it. The use of a Gaussian Mixture Model is extensively applied in such a multi-modal environment for low-level segmentation. This is trained using the expectation-maximization (EM) algorithm. The EM algorithm is used to find the maximum likelihood or maximum aposteriori features in a statistical model. It is iteratively governed by the following equations over the entire distribution.

$$\underset{\theta}{argmax}(E[log(L(\theta; X, Z))])$$

Where E is the expected value of the likelihood function L, which is given by

$$L(\theta; X) = \sum_{Z} p(X, Z|\theta)$$

X is the set of generated values from the camera; Z is the unobserved latent data and is computed from the random variable θ which is a vector that defines space where the subject is. The maximum of E over θ updates weights φ_i, mean μ_i and covariance matrix Σ_i that define the Gaussian distribution

$$p(\theta|x) = \sum_{i=1}^{K} \phi_i \eta(\mu_i, \Sigma_i)$$

The higher the probability, it is more likely to belong to the background. I thus define a threshold to allow a certain amount of pixel data under whose probability segregates the foreground subject. The cost of computational load can create a trade off on account of the approaches to this EM calculation. [19] Permuter et. al provide an

alternate approach by using k-means clustering. However, the change is negligible. OpenCV also optimizes this module to choose the former. This gives higher EM values and thus marginal improvement of performance.

3.3 Feature Extraction

3.3.1 Optical Flow

Optical flow is a method that determines apparent flow in a frame. This is usually caused by the subject in a scene. The amount of flow is calculated by the motion between two subsequent frames in a small interval say, t and t+δt. Further, the motion is in a 2D dimensional plane. Thus, the governing function is given by

$$I(x,y,t) = I(x+\delta x,\ y+\delta y,\ t+\delta t)$$

Where I is the image viewed as a frame. x,y are the mean flow in the x and y direction respectively. Since δx/δt gives velocity, applying partial differentials on the Taylor series of I over δt provides us a velocity vector at that frame which helps us better understand the spatial positions and scale of these points.

3.3.2 Lucas-Kanade Filter

Due to our previous approach of using a GMM approach to obtain clean shadow or silhouette frames, we can distinctly differentiate the subject and the scene respectively. This helps us set up the LK filter that assumes that the flow is constant in the local neighbourhood of the indexed pixel i. The image is masked the background subtracted image. This filter is defined by the local image flow velocity vector as:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \sum_i I_{x_i}^2 & \sum_i I_{x_i} I_{y_i} \\ \sum_i I_{x_i} I_{y_i} & \sum_i I_{y_i}^2 \end{bmatrix}^{-1} \begin{bmatrix} -\sum_i I_{x_i} I_{t_i} \\ -\sum_i I_{y_i} I_{t_i} \end{bmatrix}$$

Where I_x , I_y and I_t the partial derivative of image I over x,y and t respectively for each pixel in the window. Further, each vector is validated against a cartesian system categorized in intervals of /4. The vectors in each of these zones in turn contribute to information accented by the brush strokes. Within each frame, we require this window to iterate over the entire frame to compute the flow. However, if we can identify the regions where the local maxima are observed by the subject, it reduces computational load. Raptis and Soatto [17] in their paper propose to sample only over the regions of interest in the frame. This is replicated in my work since we are only concerned with the subject in the foreground. Over a sequence of frames, if no significant change is part of the subject is observed, it is given little to no weight in the computation.

We then apply this LK filter and determine the trajectory. The sampling step size of W = 20 is ideal for most datasets. Additionally, through experimental results, the eigenvalue threshold of 0.001 allowed the compromise between saliency and the sample point. This allows us to call the defined function that provides details on the margin of error per tracked point and its spatial position. From this, we can also obtain the mean, variance and angle of trajectories of optical flow.

Further, the feature vector is expanded to accommodate Histogram of Oriented Gradient (HOG) and Histogram of Oriented Optical Flow (HOOF) features. The former is derived from unsigned gradients mapping from 0 to 360 degrees and close to 9 histogram channels, after calculating the gradients. This allows the vector space to explore all degrees of freedom in motion and provide better information about the

subject's movement. Once an array of vectors is obtained, it uses the proposed concept by F. Perronnin et. al [7], [8] on Fisher vector encoding to obtain a resultant vector that best summarises the net flow in that set of frames.

So for image $I = (I_1, I_2 \ldots I_L)$ and over k dimensions, and has a Gaussian descriptor $\Phi = \{\mu_i, \Sigma_i, \pi_i; i = 1 \ldots M\}$; then the posterior probability is given by

$$p_{Ik} = \frac{e^{-\frac{1}{2}(I_i - \mu_k)^T \Sigma (I_i - \mu_k)}}{\sum_{t=1}^{M} e^{-\frac{1}{2}(I_i - \mu_t)^T \Sigma (I_i - \mu_t)}}$$

Where p_{Ik} is the posterior probability; M is the mode. After the array of optical flow vectors is obtained vectors, following which the mean and covariance vectors are given by

$$u_{jk} = \frac{1}{L\sqrt{\pi_k}} \sum_{i=1}^{L} p_{Ik} \frac{I_{ji} - \mu_{jk}}{\sigma_{jk}}$$

$$v_{jk} = \frac{1}{L\sqrt{2\pi_k}} \sum_{i=1}^{L} p_{Ik} \left[\left(\frac{I_{ji} - \mu_{jk}}{\sigma_{jk}} \right)^2 - 1 \right]$$

This 2 information are added to the feature vector space to contribute to the computation of FOA. While the descriptors extracted from here simplify the Fisher vector and approximates local image features, there are certain trivial cases which would result in 0. Thus, I continue to use the full breadth of the data to analyze non-symmetric actions.

3.3.3 Skeleton Joint System

The Kinect has a significant feature in it that allows mapping a skeleton onto the human subject in front of it. Microsoft licensed the patented concept from

PrimeSense. This uses a depth map to segment the subject and then infers the body position using a built-in computer vision module to draw the skeleton. These include depth from focus and depth from stereo principle. Shotton et. al [26] propose a new approach to quickly and accurately predict body joint points from a depth map. These concepts help get noise independent data. However, the built-in module provides sufficient accuracy with respect to screen coordinates. I can later extract the 10 body parts namely torso, head, upper/lower left/right arm/leg [24]. For better calculation, I change the reference to the coordinate system of the hip [21]. This restricts the study of movement to the subject alone. At run time, the code allows defining the mode of feature extraction we wish to adapt to perform the rest of the computation. Such an adept code helps develop a flexible system that can understand the need to work with modularized code which can be replaced and tested for further ramifications.

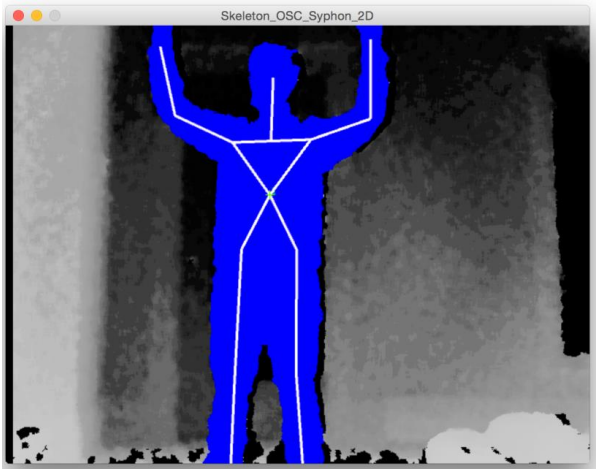

Figure 3.2. Skeleton Joint system from Microsoft Kinect

This explores the principle of isometric data. The net vectors for the approach used in my system will have a large margin of error. However, the raw data coming from the skeleton has low latency and high precision. Future work discusses ways to

incorporate skeleton data as the feature space to perform the same analysis. The number of dimensions is fixed and we are strictly concerned with the pattern of changing positions of the human to recognize and estimate their activity.

3.4 Factor Oracle

3.4.1 Overview

To delve into the construction of a sequence of frames, we extrapolate the concept of word graph into video analysis. Where depending upon the inputs sample its feature vector can best suggest a method to iterate over the linear time and space bin to find its best match in the dataset bins. The criteria to judge this frame tree utilizes the nearest neighbour search algorithm.

Allauzen et. al defines their method applied towards a word sequence. Say a word *w = abbbaab*. The Figure 3 shows how the online algorithm recognizes *aba* though it is not a factor in the test set but is a sequence that exists in the lookup database.

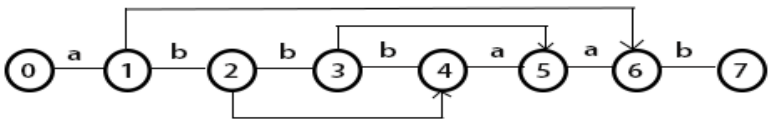

Figure 3.3. Factor Oracle matching of a sample word w

In order to identify the ideal case of sequential frames, the automaton must be

- Acyclic
- Must recognize factors of the input sequence
- Follow the path of least states; i.e. length of states recognized
- Has linear number of transitions

To elucidate, it behaves like a scrubbing window that creates a buffer to take in a set length of the feature vector, L. It looks ahead and does not loop back on the same set of frames twice. The buffer is constantly overwritten by a new set of feature values. Next, for every frame with a feature vector say $w_i \in w$ must have a factor in the lookup space. If it doesn't it can ideally skip the frame based on several constraints. They are such that, in the set of frames ahead, there is no plausible shift of state within the window, and it cannot map to the pre-indexed dataset unless there is a considerable $1+\varepsilon$ error leeway provided. Lastly, the transition from a Factor(w), which is now a string of sequence frames to recognize the state i of Oracle(w).

To understand the use of FO, it becomes clearer when viewed from the perspective of word graphs. The concept of word graph as explained by Inenaga et. al helps create a compact approach to online word-graphs [10]. To create the a word graph in real time helps facilitate the module to work without necessarily knowing the entire feature space to look in. Their drawback is that it needs to recreate the model every time a new buffer is fed into the system. The key strength of this method is its ability to create a model that works on no previous knowledge of the data coming in.

3.4.2 Priority based k-d tree

This is short for the k-dimensional tree. The most trivial definition remains that it is a binary tree where every node is a k-dimensional point. With reference to my system, it defines a feature vector $\vec{v}$. The criteria that governs the word graph decision split at each vertex is controlled by the state in FO and k-d tree optimizes the depth of each

suffix in a word. This is the most common approach of searching a space and partitioning based on the application. In this system, it uses the method as a search based error bound approach with a O(log n) search complexity. As the incoming feed of frames fills the buffer, it searches the database for a feature vector $\overrightarrow{v_i}$ that closest matches it within a margin of error say, $\vec{v} \pm \varepsilon$.

Figure 3.4. A hyperplane divide and search representation of a vector v

This process can be better understood as a hyperplane intrinsically separating the node into two. The k-means clustering reduces the search space. As the values are iteratively compared in each dimension; depending upon the constraint, it branches out into its left or right child nodes and builds the sequence as observed in word graphs or, in this case, an index that maps to the respective frames in the database. Figure 4 is an example of this hyperplane space where the vector identifies its match in the top right region within a region of frames. The varied sizes of rectangles indicate a sequence of a feature of the frames that are approximately similar. Through experimental results, it was ideal to fix ε at 0.02. As the database increases, this margin can be reduced further to incorporate a wider range of vector space. However, for the purpose of analysis, the Weizmann dataset [16] was used and provided satisfactory results. With increasing dimensions, the performance deteriorates. Additionally, this approach mapped the first occurrence of a match. For the best

match feature, however, would be considered with respect to the state of this new vector.

So, if $\overrightarrow{v_i}$ finds a Fact($\vec{v}$), at state $j > i$, the next state say k would be from $\overrightarrow{v_j}$ looking forward to suffix it to the sequence which is the Oracle($\vec{v}$). This continues until it has traversed through state i = P($\overrightarrow{v_i} \in \vec{v}, \forall\, i = 1 \ldots L$) .

Further, when a state k does not have a factor in $\vec{v}$, It merely drops the frame into another buffer that is concatenated with the new buffer on live stream. The new buffer has length L-k. In addition to these k states that are now in queue for computation.

Also, if a feature vector $\overrightarrow{v_j} \approx \overrightarrow{v_k}$ and $\overrightarrow{v_j} \approx \overrightarrow{v_m}$ where $m,k \in i$, and i is any state in the dataset of videos indexed per bin. Then the suffix is chosen based on priority. That is, if $m<k$, m is chosen as the factor. This generates a robust system to string together a sequence of frames. In most cases not all states have a factor, thus invariably becoming the shortest path through the window. The sophisticated approach suggested by [18] uses FLANN library which choses the ideal NNS method and does not consider a combinational approach to experiment with the data. Further, k-d tree query that partitions a vector space by recursively generating hyperplanes to cut along coordinates where there is maximal variance in the data. K-d tree and k-means clustering technique proposed above together is like generalised hyperplane tree (GH). There is evidence showing that k-means performs competitively with k-d tree, and that both outperform locality sensitive hashing (LSH) on real-world data. The hybrid system automatically adapts their splitting resolution according to the density of the local data and the hyperplanes used to partition the data can be along any direction, whereas in LSH the hyperplanes are constrained to align with coordinate directions.

3.4.3 Jaro-Winkler Distance

Since k-d trees iterate over the feature space one dimension at a time, it is proving costly over every suffix but is reasonable on the entire system. Thus, an alternate approach is needed to allow a marginal region of acceptance that not only looks ahead, but also matches with a vector that is above a certain threshold of similarity in that chosen sequence. To quantify the threshold, I use a measure of similarity conventional in major statistical problems using the Jaro-Winkler distance [5]. This helps consider the breadth of each vector. It is an iterative process as it adds a suffix of a new dimension. So as k-d tree performs its iteration, another thread calculates the distance of the vector of *m*-dimensions, for $m<k$. The J-W distance d_w is given by:

$$d_w = d_j + (lp(1 - d_j))$$

Where *l* is the length of the prefix in the sequence before you calculate the distance. *p* is the scaling factor. Through prior knowledge fixed at 0.1 and d_j is given by

$$d_j = \begin{cases} 0 & if m = 0 \\ \frac{1}{3}\left(\frac{m}{|\vec{v_1}|} + \frac{m}{|\vec{v_2}|} + \frac{m-t}{m}\right) & else \end{cases}$$

Where *m* is the number of matching dimensions, *t* is half the number of transpositions and $\overrightarrow{v_1}$, $\overrightarrow{v_2}$ are the length of the dimensions considered. This approach provides a fuzzy matching of feature space. Thus at any point there is a Fact($\vec{v}$) such that, when matched is very similar to the frame analyzed.

3.4.3 1 versus all non-linear SVM

The open source machine-learning library LIBSVM implements the algorithm for kernel SVM. SVM requires data to be represented as a vector of real numbers. The

first step was transforming the data into numerical data and then to the format for the LIBSVM package. While choosing a model for the SVM, several parameters are taken into account such as the penalty parameter, C and the kernel parameters. I found that the model worked best when the soft margin constant C was kept at 100. The smaller value of C will tend to ignore the points close to the boundary and causes false results. Kernel parameters also have a significant effect on the SVM model. As the feature set is small, I chose the RBF kernel as it non-linearly maps data into a higher dimensional space and handles non-linear relationships between class labels and features. The degree of the polynomial controls the flexibility of the classifier. I found that the 5- degree polynomial works best as it has a greater curvature. The nu-SVM model sets a lower and upper boundary on the number of data points that lie on the wrong side of the hyperplane and is advantageous for controlling the number of support vectors.

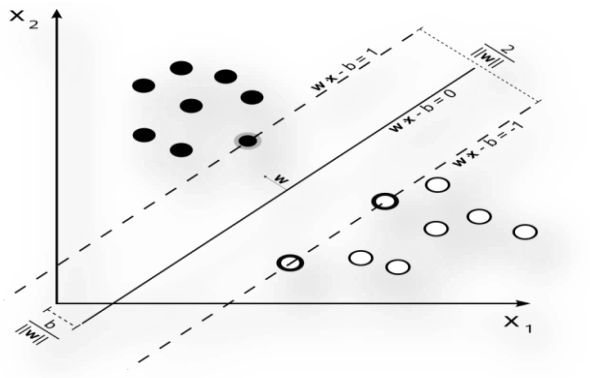

Figure 3.5. A linear SVM model to represent separating hyperplane

Where,

$$w.\mathbf{x} - b = 1$$

$$w.\mathbf{x} - b = -1$$

Are the governing equations; $\vec{w}$ is the normal vector to the hyperplane, x is the feature vector and $\frac{b}{\|\vec{w}\|}$ determines the offset of the hyperplane.

The RBF kernel is given by,

$$K(x, x') = e^{-\gamma \|x - x'\|^2}$$

Where $\gamma = 1/2\sigma^2$; x and x' are two samples, σ is a free parameter.

This feature only exists to understand and test the model for its accuracy of prediction and estimation of human activity. The primary belief about real-time analysis is not to classify but to append the sequence.

CHAPTER 4: RESULTS

4.1 The Offline Dataset

The Weizmann dataset [16] was used. For the prototype, the study uses these actions to level against the online dataset. The base of the algorithm estimates from this dataset. As the dataset grows, it creates better estimates to the analyzed human activity is played back.

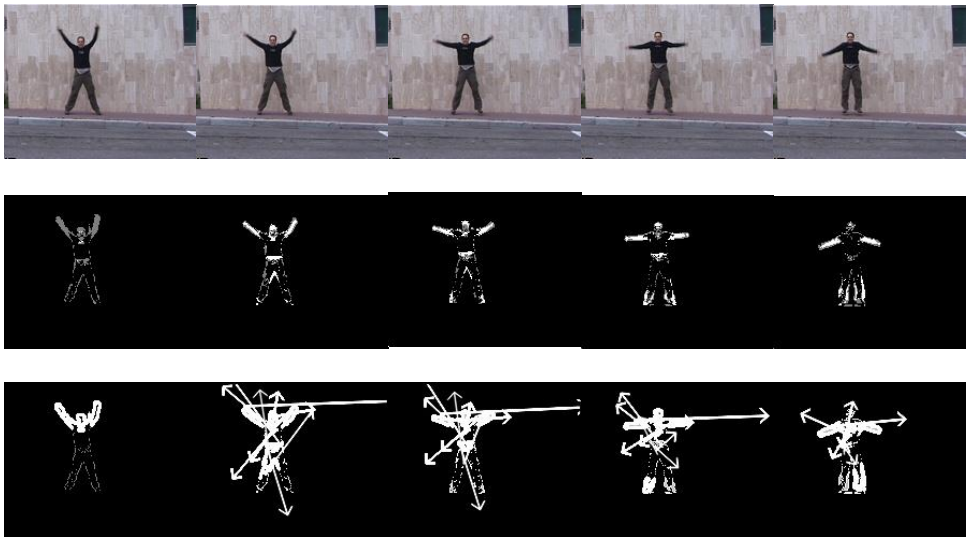

Figure 4.1. Set of dataset frames as viewed from a (a)-(e) raw feed, (f)-(j) Segmented and (k)-(o)feature extracted perspective.

Pearson's correlation analyzed on various extracted attributes that constitute the dataset can be visually perceived as

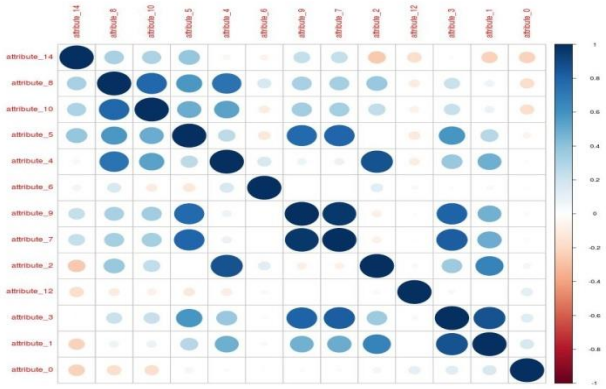

Figure 4.2. Pearson's correlation of 16 attributes

The percentage of correlation is shown by the shade of the circle. Closer to red implies inverse proportionality to correlation and vice versa.

4.2 The Online Dataset

The proposed system allows creating bins from the live stream. This is done in real time and usually creates a frame length defined at run-time.

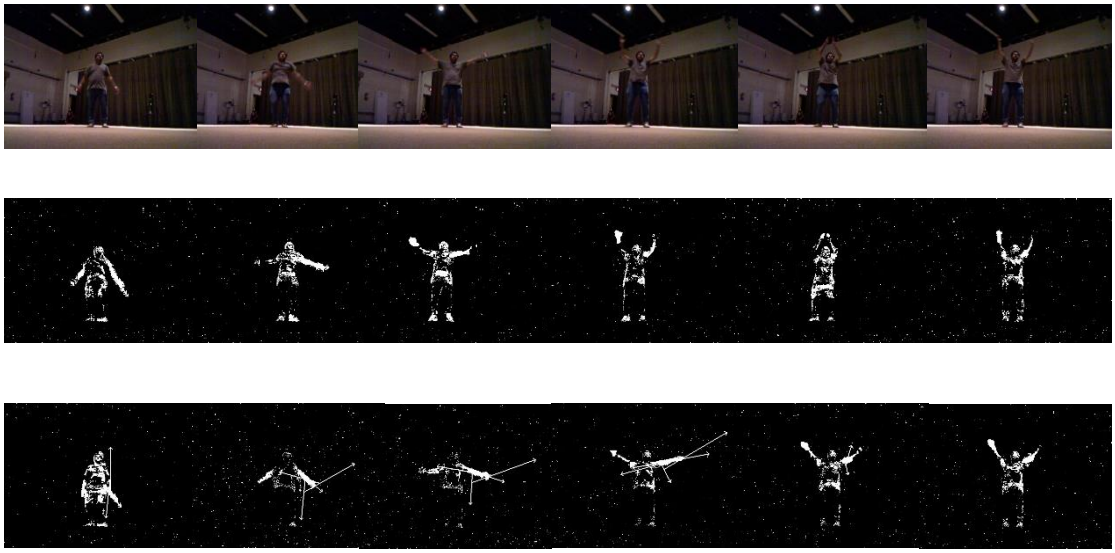

Figure 4.3. Set of input frames as viewed from a (a)-(e) raw feed, (f)-(j) Segmented and (k)-(o)feature extracted perspective.

4.3 Pattern Retrieval

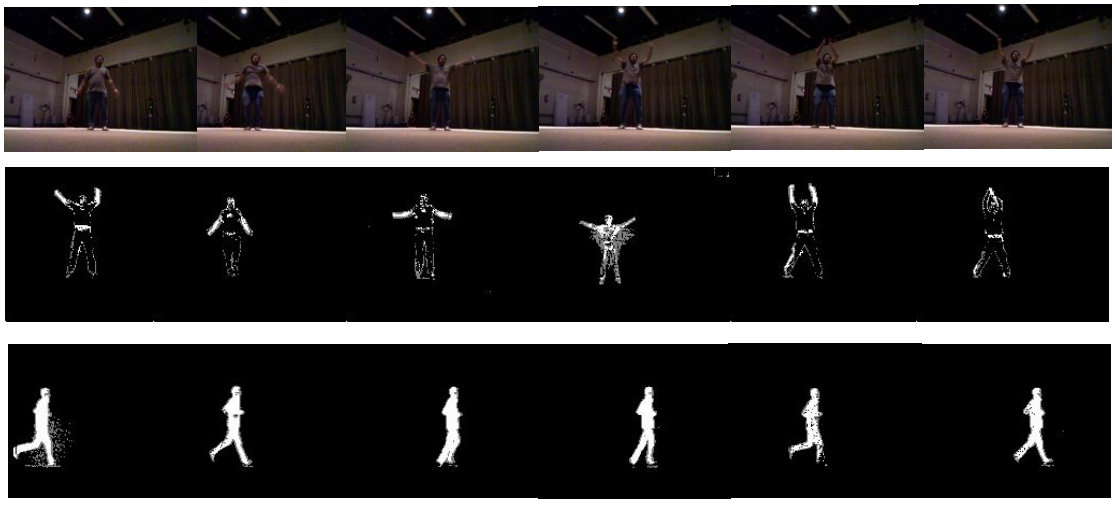

Figure 4.4. Set of input frames as viewed from a (a)-(e) raw feed, (f)-(j) first order pattern retrieval and (k)-(o)second order predicted pattern match

As an action is recognized it splits the playback region into a 1st order similarity of actions or patterns. Further, it also predicts the likelihood of the 2nd order set of frames that could, on priority hold contextual inference from the primary pattern and look ahead into a plausible momentary future.

4.4 K-means clustering

This was a method proposed by [18] where the reduced search space helps better the speed to search the nearest feature pattern that matched this cluster. When the data was segregated at run time over 50 iterations had the following result.

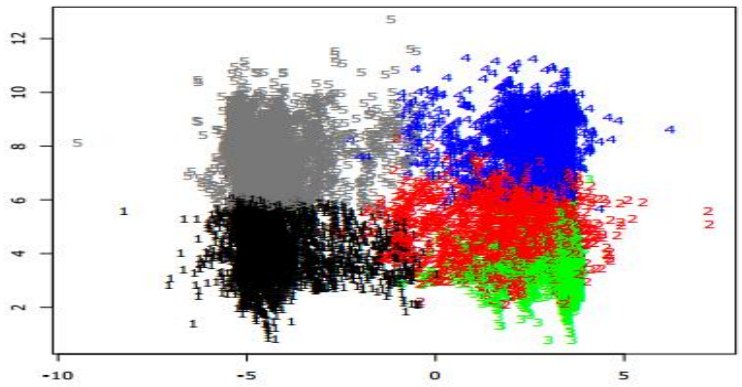

Figure 4.5. K-means clustering on 5 actions from offline dataset

Clearly there is no saturation of the clusters and there are several outliers. However, within a considerable margin of error, the search space-time to find the boundary improved significantly. The model took 0.34seconds to build. There is definitely a toll on performance due to the curse of dimensionality.

4.5 Randomized k-d tree

The system used a multi-thread process to run in parallel and analyze multiple randomized k-d trees. Where the number of dimensions was 16 and a tree depth of 12

was chosen. Beyond which it reached saturation on the performance curve. It took 0.84 seconds to build the model without k-means and 0.68seconds with k-means. Further over the Weizmann dataset of 5 actions each of 10 bins, a 20,494 sample space was obtained.

To test the accuracy of the model, when tested as a classifier, it pitched a 92.406%. Further, this model had a mean absolute error of 0.03 and RMSE of 0.155.

1	2	3	4	5	Action
6306	27	34	34	27	1
21	2263	2	8	9	2
30	1	3023	291	303	3
31	7	444	3569	201	4
25	6	122	61	5333	5

Table 1. Confusion Matrix k-d Tree

4.6 Jaro-Winkler Probability distribution

This approach performs much faster. It takes 0.26 seconds to build the model. So, when coupled with the k-d tree, it is still under a second and is palatable as real-time to the user. To understand how the distance varies per epoch of buffer over-written; a hybrid model with an error bound k-d tree is proposed. The average is 0.9. This can be visually seen in Figure below.

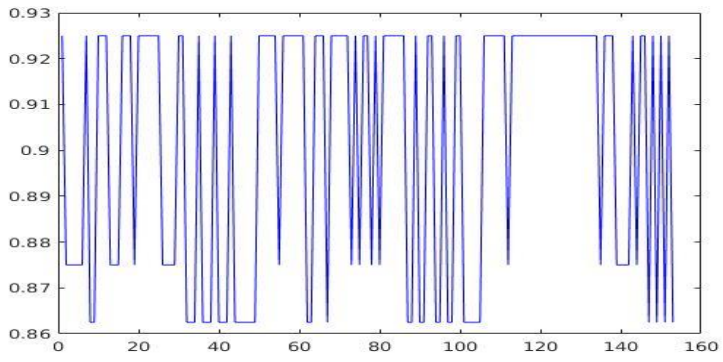

Figure 4.6. Distance range of Jaro-Winkler distance over estimated frames

4.7 1 versus all SVM Output

The feature vector spanned over 16 dimensions. Further, when each bin was segregated at run time before the SVM analysis, the average of each dimension of each subset of each bin was computed as a feature vector. This was then used to lazy train the SVM model and the obtained accuracy was 94.472%. It could correctly classify a class and it was run by the other classes to get the winning class per vector. It took 0.89 seconds to build the model. Further, the model's RMSE was approximately 0.12.

1	2	3	4	5	Action
2166	18	4	8	2	1
3	771	2	39	3	2
3	0	1156	49	18	3
3	0	79	1317	9	4
0	0	7	19	1865	5

Table 2. Confusion Matrix SVM

CHAPTER 5: FUTURE WORK

In this thesis, I have presented a framework that addresses a new perspective on the application of Factor Oracle algorithm. This has potential applications is a wide range of industry. This has the ability to evolve into an unsupervised pattern learning module that could perform high detailed animation. It also provides a platform to understand the underlying architecture involved in activity recognition. These underlying concepts are: acquiring the data, pre-processing it, segmentation, feature extraction, algorithmic perception of the data and driving the output towards meaningful results. In this system, it estimates the activity using Optical Flow and recommends a sequence of frames that best match the action.

I proposed a solution based on dynamical analysis via recurrence relations, which has an interpretation in terms of geometric structures of high-dimensional data. This framework also shows the ability to compress this feature space to contain discrete information that helps to still create high-level computation within a small domain. However due to the curse of dimensionality and the envelope of complexity-time trade off in proportion to the capability of the system, opens up the possibility of a big data computation using Master-Slave architecture.

This phenomenon allows the system to accommodate a larger dataset and also incorporate the use of MapReduce algorithm with Apache Hadoop. In this, the master node has several worker nodes that scan the database simultaneously for effective pattern matching per incoming frame. This will significantly drop the processing time faster and allows a heavier algorithm to process data in high dimensionality.

This creates opportunities to multi-threaded the system and in parallel create multiple pattern approaches like the FLANN library and improve the framework stability. This will also improve accuracy towards human activity estimation and recognition.

Further, due to its high sensitivity to ambient lighting, it becomes interesting to test the system and push it to different experiential controls. The expansion on this system is to ensure its application in a production based environment where it creates a full circle by automating shadow puppets through such a predictive system. Thus, enabling the performer to choreograph a sequence that estimates a contextual story out of the sequence of movements. Additionally, the end user can derive useful results by extrapolating this to self - driving cars where it can perform predictive analysis on obstacles ahead of it.

REFERENCES

[1] Andoni and P. Indyk, “Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions,” Commun. ACM, vol. 51, no. 1, pp. 117–122, 2008.

[2] Aizerman, Mark A.; Braverman, Emmanuel M.; and Rozonoer, Lev I. (1964). "Theoretical foundations of the potential function method in pattern recognition learning". Automation and Remote Control 25: 821–837.

[3] Allauzen, C; Crochemore M; Raffinot, M (1999), Factor oracle: a new structure for pattern matching, Lecture Notes in Computer Science, Vol 1725, p.1-16, ISBN 3-540-66694

[4] Assayag, G; Bloch, G; Chemillier, M; Cont, A; Dubnov, Shlomo, (2006), ACM Multimedia Workshop on Audio and Music Computing for Multimedia, ISBN 1595935002

[5] Cohen, William W; Fienberg, Stephen E; Ravikumar, Pradeep D; Fienberg, Stephen E, 2003, A Comparison of String Distance Metrics for Name-Matching Tasks, Proceedings of IJCAI-03 Workshop on Information Integration on the Web, p. 73-78

[6] Danciu, Gabriel; Banu, SM; Caliman, A, (2012), International Conference on System Theory, Control and Computing (ICSTCC), IEEE, ISBN 978-1-4673-4534-7

[7] F. Perronnin and C. Dance. Fisher kenrels on visual vocabularies for image categorizaton. In Proc. CVPR, 2006.

[8] Florent Perronnin, Jorge Sánchez, and Thomas Mensink. Improving the fisher kernel for large-scale image classification. In *Proc. ECCV*, 2010.

[9] H.R. Maturana, F.J. Varela (1980). "The cognitive process". Autopoiesis and cognition: The realization of the living. Springer Science & Business Media. ISBN 978-9-027-71016-1.

[10] Inenaga, Shunsuke; Hoshino, Hiromasa; Shinohara, Ayumi; Takeda, Masayuki; Arikawa, Setsuo; Mauri, Giancarlo; Pavesi, Giulio, 2005, On-line construction of compact directed acyclic word graphs, Discrete Applied Mathematics, Vol. 146, Issue 2, p. 156-179

[11] Ke, Shian Ru; Thuc, Hoang; Lee, Yong Jin; Hwang, Jenq Neng; Yoo, Jang Hee; Choi, Kyoung Ho, (2013) A Review on Video-Based Human Activity Recognition, MDPI Computers Journal, p.88-131, ISBN 1206708344

[12] Kim, Eunju; Helal, S; Cook, D, 2010, Human Activity Recognition and Pattern Discovery, Pervasive Computing, IEEE, Vol.9, Issue 1, p.48-53, ISBN 1536-1268

[13] Kui Jia and Dit-Yan Yeung. Human action recognition using Local Spatio-Temporal Discriminant Embedding. In CVPR 2008.

[14] Kurniawati, Ruth; Jin, Jesse S; Shepherd, John A, 1998, Efficient Nearest-Neighbour Searches Using Weighted Euclidean Metrics, Advances in Databases, p.64

[15] Laufer, Berthold, 'Oriental theatricals- Part 1', in Field Museum of Natural History, Chicago, 1923 (The University of Michigan, May 21, 2008), Page 36.

[16] Lena Gorelick and Moshe Blank and Eli Shechtman and Michal Irani and Ronen Basri,2007, Actions as Space-Time Shapes, Transactions on Pattern Analysis and Machine Intelligence, Vol 29, Issue 12, p.2247-2253

[17] M. Raptis and S. Soatto, "Tracklet descriptors for action modeling and video analysis," in European Conference on Computer Vision, 2010

[18] Muja, Marius; Lowe, David G, 2014, Scalable Nearest Neighbour Algorithms for High Dimensional Data, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 36, Issue 11, p. 2227-2240

[19] Permuter, H.; Francos, J.; Jermyn, I. A study of Gaussian mixture models of color and texture features for image classification and segmentation. Pattern Recogn. 2006, 39, 695–706

[20] Raviteja Vemulapalli; Felipe Arrate; Rama Chellappa, 2014, Human Action Recognition by Representing 3D Skeletons as Points in a Lie Group, CVPR

[21] Reeves, William T; Salesin, David H; Cook, Robert L, (1987), Rendering antialiased shadows with depth maps, ACM SIGGRAPH Computer Graphics, Vol 21, Issue 4, p.283-291, ISBN 0897912276

[22] S. Ali, A. Basharat, and M. Shah. Chaotic invariants for human action recognition. In Proc ICCV, 2007.

[23] S. Arya, D. M. Mount, N. S. Netanyahu, R. Silverman, and A. Y. Wu, “An optimal algorithm for approximate nearest neighbor searching in fixed dimensions,” J. ACM, vol. 45, no. 6, pp. 891– 923

[24] Sedai, S.; Bennamoun, M.; Huynh, D. Context-based Appearance Descriptor for 3D Human Pose Estimation from Monocular Images. In Proceedings of IEEE Digital Image Computing: Techniques and Applications (DICTA), Melbourne, VIC, Australia, 1–3 December 2009; pp. 484–491

[25] Sen, Pradeep; Cammarano, Mike; Hanrahan, Pat (2003), Shadow silhouette maps, ACM Transactions on Graphics, Vol 22, Issue 3, p.521, ISBN 1-58113-709-5

[26] Shotton, Jamie; Fitzgibbon, Andrew; Cook, Mat; Sharp, Toby; Finocchio, Mark; Moore, Richard; Kipman, Alex; Blake, Andrew, 2011, Real-time human pose recognition in parts from single depth images, CVPR, p. 1297-1304

[27] Sundaram, Narayanan; Brox, Thomas; Keutzer, Kurt, 2010, Dense point trajectories by GPU-accelerated large displacement optical flow, Lecture Notes in Computer Science, Vol.6311 LNCS, Issue 1, p.438-451, ISBN 3642155480

[28] V. N. Edward, (1980) Javanese Wayang Kulit: An Introduction: Oxford University Press.

[29] W. Huag, Q. Wu, 2010, Human Action Recognition using Recursive Self Organizing Map and Longest Common Subsequence Matching, IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), , p. 2130-2133

[30] Wang, Heng; Kläser, Alexander; Schmid, Cordelia; Liu, Cheng Lin, 2013, Dense trajectories and motion boundary descriptors for action recognition, International Journal of Computer Vision, Vol.103, Issue 1, p.60-79, ISBN 1126301205

[31] Wittgenstein, Ludwig (2001) [1953]. “Philosophical Investigations”. Blackwell Publishing. p.6. ISBN 0-631-23127-7.

[32] Wittgenstein, Ludwig (2001) [1953]. “Philosophical Investigations”. Blackwell Publishing. p.9

ABOUT THE AUTHOR

Prashanth is a Theatre enthusiast, keyboardist and trained contemporary dancer, pursuing his MS in EE+AME. His research focuses on amalgamation of engineering and arts predominantly in the field of experiential systems and augmented reality.

Prashanth has his roots in electrical engineering. With interests in signal processing, through this he aims at analysing both from its front and back end, the structure of a system and its organizational behaviour. His work is founded upon media installations for motion capture and gauging quality of gestures. Focussing on image and video processing, Prashanth hopes to delve into high resolution projection of published content and pave a future towards multi-dimensional perception of the environment.

Prior to attending ASU, he worked with Samsung R&D - India on the smart TVs in its Graphic User Interface Domain. In addition to this worked on its Cloud Computing Department. He designed a middleware architecture using Machine Learning through Apache Libraries to deploy a text semantics application for unstructured data. This laid the stepping stones for his interest in organizational level coding through Enterprise Architecture. He earned his Bachelor's of Technology majoring in Electrical Engineering from SASTRA University, India.