MULTIMODAL FEATURE ANALYSIS FOR QUANTITATIVE PERFORMANCE EVALUATION OF ENDOTRACHEAL INTUBATION (ETI)

Samarjit Das[1], *Jestin Carlson*[2], *Fernando De la Torre*[1] *and Jessica Hodgins*[1]

[1]The Robotics Institute, Carnegie Mellon University, Pittsburgh PA 15213
[2]Department of Emergency Medicine, University of Pittsburgh, Pittsburgh PA 15260

ABSTRACT

Endotracheal intubation (ETI) is a very crucial medical procedure performed on critically ill patients. It involves insertion of a breathing tube into the trachea i.e. the windpipe connecting the larynx and the lungs. Often, this procedure is performed by the paramedics (aka providers) under challenging prehospital settings e.g. roadside, ambulances or helicopters. Successful intubations could be lifesaving, whereas, failed intubation could potentially be fatal. Under prehospital environments, ETI success rates among the paramedics are surprisingly low and this necessitates better training and performance evaluation of ETI skills. Currently, few objective metrics exist to quantify the differences in ETI techniques between providers of various skill levels. In this pilot study, we develop a quantitative framework for discriminating the kinematic characteristics of providers with different experience levels. The system utilizes statistical analysis on spatio-temporal multimodal features extracted from optical motion capture, accelerometers and electromyography (EMG) sensors. Our experiments involved three individuals performing intubations on a dummy, each with different levels of expertise. Quantitative performance analysis on multimodal features revealed distinctive differences among different skill levels. In future work, the feedback from these analysis could potentially be harnessed towards enhancing ETI training.

Index Terms— Multimodal feature analysis, 3D landmark shape, EMG, endotracheal intubation, emergency medicine

1. INTRODUCTION

Endotracheal intubation (ETI), or insertion of a breathing tube into the trachea (the windpipe connecting the larynx and the lungs), is a very crucial life-saving procedure performed on critically ill patients under emergency settings [1]. Successful intubations provide an unobstructed airway to the lungs of the patients under trauma and/or severe injuries who are unable to breathe otherwise. Most often, this procedure is performed by the paramedics under challenging prehospital settings and very critical time constraints. In those situations, the timing and outcome of intubation attempts can be the deciding factor between life and death. Unsuccessful intubation and/or a delay in performing the task can often lead to fatal consequences. Given its critical importance, it is quite surprising and alarming that the success rate of ETI among paramedics in the US is as low as 45% [2, 3, 4]. Thus it is it is of utmost importance to efficiently train and evaluate the task performances of the paramedics providing this crucial emergency service.

Endotracheal intubation is performed with the help of a laryngoscope, an instrument with a curved blade for visualization of the glottis i.e. the front door to the trachea. During intubation, the paramedics insert the laryngoscope into the mouth for a direct visualization of the glottis before inserting the breathing tube into it. This is a very crucial step and delicate maneuvers with the laryngoscope is necessary for a successful intubation. Few objective metrics exist to quantify the differences in ETI techniques between providers of various skill levels. Current evaluation schemes including binary success/failure methods [5] or video laryngoscopy based analysis [6] are inadequate for providing impactful feedback towards training as well as lack the ability to objectively track the learning curve of a intubation trainee. In this pilot study, we develop a quantitative framework for discriminating the kinematic characteristics of providers with different experience levels. The system utilizes statistical analysis on spatio-temporal multimodal features extracted from optical motion capture, accelerometers and electromyography (EMG) sensors while the providers attempted to intubate a dummy. It is to be noted that motion capture has previously been used for evaluating medical procedures e.g. ocular [7] and laparoscopic surgery [8].

Motion capture enabled us to track the 3D movements of the provider's upper body, the dummy and the laryngoscope simultaneously. We parameterized the dynamics of arm/head movements w.r.t the dummy using a piece-wise stationary 3D landmark shape deformation model (see Sec. 2.1.1 for details). The basic idea is to capture the distinctive pattern of motor movements during intubation as a sequence of deforming landmark shapes [9, 10]. Our analysis was motivated by the fact that different skill levels would be reflected in the temporal movement patterns and thus leading to discriminative shape dynamical parameters. Apart from shape deformation features, we also computed spatio-temporal features from the EMG and accelerometer sensors placed at the wrist and biceps of the providers. These include DC level, power/energy, spectral entropy and cross-correlation [11]. Several features were also extracted from the 3D orientation profile of the laryngoscope during ETI.

Our experiments involved three individuals; each with different levels of expertise, performing multiple trials of ETI. Quantitative performance analysis on multimodal features revealed distinctive differences among different skill levels. Further work in this direction might have useful impact on more objective evaluation and training of ETI as well as any other job-coaching application involving skillful motor movements.

2. METHODS

In this section, we first discuss our multi-sensor system setup and then move on to discuss the feature computation frameworks.

2.1. System Description

The intubations were performed on a mannequin (i.e. dummy) using a laryngoscope as shown in Fig. 1. The laryngoscope was equipped

Thanks to NSF for supporting this project.

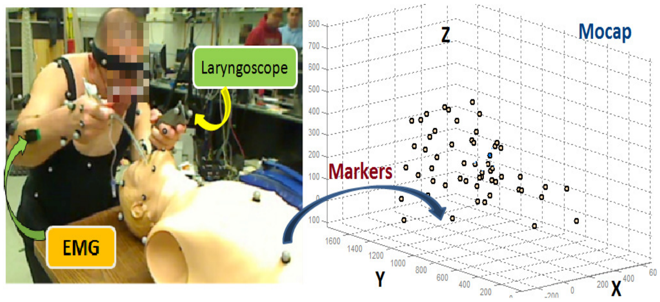


Fig. 1. This figure demonstrates the experimental setup. The providers are fitted with reflective markers (white spheres) as well as EMG/accelerometer sensors while they performed intubation on a dummy. The 3D marker trajectories of all the markers were tracked during the entire procedure (one frame is shown on the right).

with a video camera that enabled us to verify the successful placement of the breathing tube. For motion capture, we used a Vicon [12] system with 16 near-infrared cameras. Each capable of recording 4 megapixel resolution images at 120 Hz. A total of 50 retro-reflective markers were placed on the upper body of the subjects, another 18 were placed on the dummy and 3 were placed on the laryngoscope handle. Motion capture facilitated highly accurate tracking of the 3D trajectories corresponding to all the markers during intubation. This enabled us to accurately record the movements of the body, arms and head of the subjects, 3D orientation of the laryngoscope and their movements w.r.t the dummy. We also placed EMG sensors at various muscle locations of the arms in order to track the spatio-temporal muscle activation patterns during intubation. Each of these sensors were also equipped with a triaxial accelerometer that could record abrupt motion variations/jerks associated with the arms. All EMG/accelerometer time-series data were transmitted wirelessly to a hub and they were synchronized with the motion capture data for joint multimodal feature analysis.

2.2. Multimodal Feature Extraction

First, we discuss the 3D landmark shape deformation framework to model the temporal dynamics associated with the movements of arm/head of the subjects w.r.t the dummy. Then we explain the feature extraction techniques for laryngoscope orientation variation profile as well as EMG/accelerometer data. All these features were computed over overlapping temporal windows during ETI. The intubation time interval is defined from the instant the laryngoscope is inserted into the mouth to the instant when it is removed following the placement of the tube.

2.2.1. 3D Landmark Shape Deformation Model

A 3D landmark shape is represented by an ordered set of points (aka landmarks) in the 3-dimensional space [9]. Thus, a k-dimensional landmark shape is represented by a $k \times 3$ matrix with each row containing the x,y,z coordinates of the corresponding point. In our case, the markers are treated as landmark points. At each instant, we represent the collective 3D locations of the head and the left arm[1] together with the dummy's mouth as a 8-dimensional 3D landmark shape as shown in Fig. 2. Now, temporal movement patterns associated with the head and left arm w.r.t the dummy's mouth can be modeled as a deforming landmark shape sequence (see Fig. 2). The corresponding feature vector is computed as the parameters of shape deformation dynamics over the window. Our goal is to capture the distinctive pattern of motor movements associated with different ETI skill levels in terms of the shape dynamical parameters.

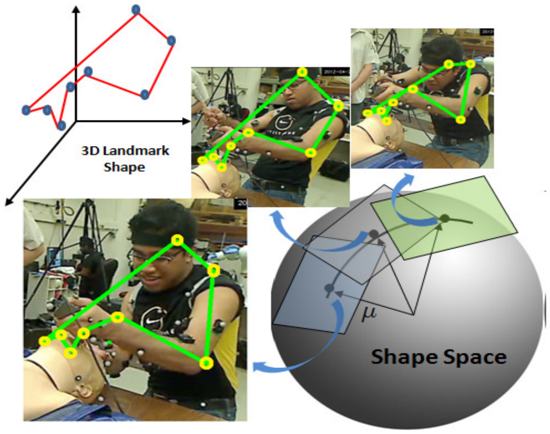


Fig. 2. This figure demonstrates how we represent the temporal movement patterns of left arm/head w.r.t the dummy's mouth as a sequence of deforming 3D landmark shapes.

For modeling the shape deformations, we use a piece-wise stationary shape activity (PSSA) model. This is similar in spirit to Non-stationary Shape Activity or NSSA model (see [10] for details). The difference is, in PSSA, we define a single procrustes mean shape, μ [9], over the temporal window and model the shape deformation dynamics w.r.t it (see Fig. 2). In other words, we consider a stationary shape sequence within each window, and hence the name *piecewise stationary* model. This is unlike NSSA, where the mean shape changes at each time instant [10]. The PSSA shape deformation feature computation framework goes as follows.

Say, we have N time frames in the temporal window. The corresponding locations of the subject's head, left arm and the dummy's mouth is represented by the sequence $\{S_t\}_{t=1}^{N} : S_t \in \mathcal{R}^{8\times3}$ (see Fig. 2). As shown in [9], we compute the procrustes mean shape μ from $\{S_t\}$ and define a tangent space U w.r.t μ in the shape space. The columns of U contain the orthonormal basis set that spans the tangent space. Now at each t, S_t is scale, translation and rotation normalized w.r.t. μ to compute the landmark shape z_t aligned to the local mean shape (details in [9, 10]). The relative shape deformation w.r.t. the mean shape i.e. $z_t - \mu$ is then projected to the tangent space to compute v_t. The temporal shape deformation characteristic over the window is modeled by learning the parameters governing the time evolution of v_t i.e. we fit parametric time series model to $\{v_t\}_{t=1}^{N}$. The time-series parameters gives the feature vector over the current window. This process is repeated for each temporal window. The entire procedure is summarized in Algorithm 1.

2.3. Other Features

For a successful intubation, it is crucial to perform the right maneuvers with the laryngoscope. Hence, we compute several features associated with the orientation variations of the laryngoscope during ETI. These are computed from the 3D trajectories of the markers placed on the laryngoscope. One such feature is the laryngoscopic plane (LP) feature. We denote it as $LP(\theta_t)$ which is defined as: $LP(\theta_t) = \sqrt{\theta_x^2 + \theta_y^2 + \theta_z^2}$ where $\theta_x, \theta_y, \theta_z$ are the angles at time t, made by the normal to plane formed by the three markers placed on the laryngoscope (front, side and back of the handle). The temporal variations of $LP(\theta_t)$ can be used to characterize various steps in the intubation process. A typical profile of $LP(\theta_t)$ is plotted for a single intubation attempt in Fig. 3. Notice that the ETI time interval

[1] The left arm performs crucial maneuvers with the laryngoscope and its dynamics is of particular interest to us.

Algorithm 1 Computation of Shape Deformation Features

Input: $\{S_t\}_{t=1}^N : S_t \in \mathcal{R}^{k\times 3}$ (landmark configurations, $k = 8$)
Output: 3D landmark shape deformation parameters

First, compute the mean shape $\mu = \mu(S_1, ..., S_N)$ and the tangent space basis set $U = U(\mu)$ as shown in [9, 10]. Then do:

1. For each t, compute $z_t = w_t \mathcal{U}\mathcal{V}^T$ where, $\mathcal{V}\Lambda\mathcal{U}^T = \text{SVD}(\mu^T w_t)$. The term SVD means singular value decomposition and w_t is the scale-translation normalized version of S_t [9]
2. For each t, compute tangent space projection of shape deviation $v_t(z_t, \mu) = [I_{3k} - vec(\mu)vec(\mu)^T]vec(z_t)$. Here, $vec()$ denotes vectorization operation. I denotes identity matrix.
3. Compute tangent space projection coefficients $\{c_t\}_{t=1}^N$ from $\{v_t\}_{t=1}^N$ as $c_t = U^T v_t$
4. Compute B by performing Principal Component Analysis or PCA on $\{c_t\}_{t=1}^N$ and then compute $\{p_t\}_{t=1}^N$ with $p_t = B^T c_t$ where $p_t \in \mathcal{R}^d, d << N$ and columns of B contains top d eigen vectors of the covariance matrix of $\{c_t\}$ [13]
5. Learn AR(1) model $[A \;\; \Sigma]$ with $p_t = Ap_{t-1} + n_t$, $n_t \sim \mathcal{N}(0, \Sigma)$ from $\{p_t\}_{t=1}^N$ where $\mathcal{N}(\mathbf{0}, \Sigma)$ denotes a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix Σ.
6. Output feature vectors as: $Diag(\Sigma) = [\sigma_1^2, \sigma_2^2 \ldots, \sigma_d^2]^T$

can be split into four distinct zones: A, B, C and D. The temporal information of these four zones would be used to localize the discriminative features so that the training procedure can focus more on the corresponding step. Other features include the mean and variance of LP angular speed $\propto LP(\theta_t) - LP(\theta_{t-1})$, spectral entropy of $LP(\theta_t)$ and mean/variances associated with vertical movements of the LP.

The features for EMG and accelerometer data were only computed at the wrist and biceps locations. These features, computed over the temporal windows for each location, include : mean (DC) level, power and spectral entropy [11]. We also computed the temporal cross-correlation between EMG muscle activation signals from the biceps and the wrist (left arm). We hypothesize that the key motor movements associated with different skill levels will leave some signature on the corresponding spatio-temporal muscle activation patterns. In the next section, we demonstrate visual as well as quantitative comparisons of the multimodal features across individuals with different skill levels in ETI.

3. EXPERIMENTS AND FEATURE ANALYSIS RESULTS

Our experiments involved three subjects with different levels of experiences in ETI - one experienced (attending physician), one intermediate (resident in Emergency Medicine) and one novice provider (with no previous ETI experience; went through a quick training session prior to the experiments). Each performed four ETI attempts on a dummy. All twelve trials were successful as reviewed by the laryngoscope video. The mean durations of ETI attempt (std. deviation) in seconds were 5.50 (0.68) for the experienced, 6.32 (1.13) for the intermediate and 12.38 (1.06) for the novice provider.

Next, we compared the temporal variation characteristics of the shape deformation features (Sec. 2.1.1) over all the trials across the three subjects. The profile of σ_1^2 over all the trials is compared for the experienced, intermediate and the novice in Fig. 4a. It can be clearly seen that the repetitive pattern occurring at the end of zone C for the experienced reduces in amplitude for the intermediate and finally diminishes in case of the novice. Similar variation profiles were also observed for other components of Σ. This signifies a key maneuver at the time of tube placement (zone C terminus) predominantly associated with the experienced provider. In Fig. 4b, we compare the cross-correlation feature between EMG signals from the left wrist and biceps. The novice clearly has more zero crossings which can be attributed to the frequent changes in untrained spurious arm movement patterns. We also compared the variance profiles of LP angular speed (Sec. 2.3) in Fig. 4c. The modes of the curve corresponds to small back and forth movements with the laryngoscope. More the number of modes, more is the extent of redundant movements. The experienced is found to have the least of these movements whereas, the novice has been found to have the most. Next, we also compared the muscle activation patterns using EMG signals from the left wrist as well as the inter-trial variability of the laryngoscopic orientation profile i.e. $LP(\theta_t)$ (see Fig. 5). The relative quantitative comparisons across various features are shown in Fig. 6. It is to be noted that we omit the discussion of features that did not have significant discrimination across skill levels (e.g. accelerometer features).

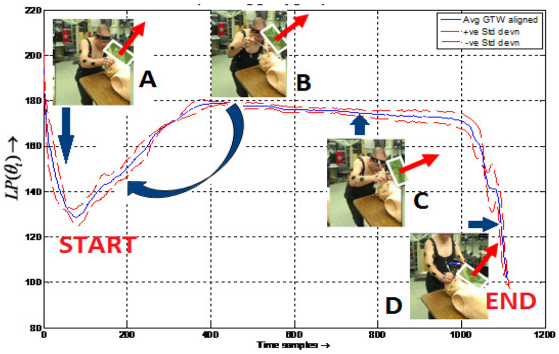
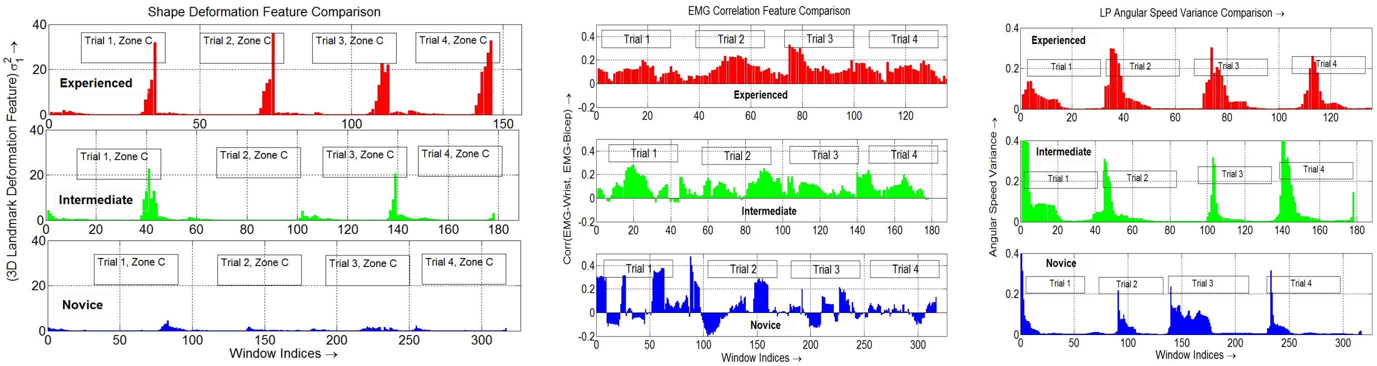


Fig. 3. Various time zones during ETI. The first down sloping portion of the curve corresponds to the laryngoscope entering the mouth (A). The upslope corresponds to obtaining the view of the vocal cords (B). The plateau corresponds to holding the view constant while placing the breathing tube (C). The final downslope represents removal of the laryngoscope after successfully placing the tube (D). The red arrows indicate laryngoscope orentation.

Finally, we performed PCA on the multimodal feature space (see Fig. 7a) and computed the distance $D(.)$ among the cluster centers corresponding to different subjects in a 3-dimensional subspace. It was found to be: D(Experienced, Intermediate) = 12.40, D(Experienced, Novice) = 151.67 and D(Intermediate, Novice) = 150. Further, in order to compare the discriminative aspects of the features over various zones of ETI, we performed a k-means ($k = 10$) clustering over all the feature data points and represented various zones as a histogram of associated cluster centers. The comparison results indicate that zone A and B are the most discriminative across skill levels (i.e. the maneuvers starting from laryngoscope insertion and glottis visualization). The histogram representation for zone A and B is shown in Fig. 7b. Observe the similarity of representations for the experienced the intermediate provider.

In conclusion, the quantitative performance analysis on multimodal features revealed distinctive differences among different skill levels. We were also able to localize some of the differences across various zone of ETI. The feedback from these analysis could be used for more objective performance evaluation as well as improving ETI training.



a) Shape deformation feature　b) Cross-Correlation feature (EMG)　c) Variance of LP angular speed

Fig. 4. Fig. a compares σ_1^2 variation profile across subjects (the four trials are cascaded together, window length 240, slide length 20). Notice that the distinctive repetitive patterns associated with the experienced provider. Fig. b and Fig. c compares the EMG cross-correlation of wrist-biceps and variance profile of LP angular speed respectively. The significance of these discriminative patterns are discussed in Sec. 3

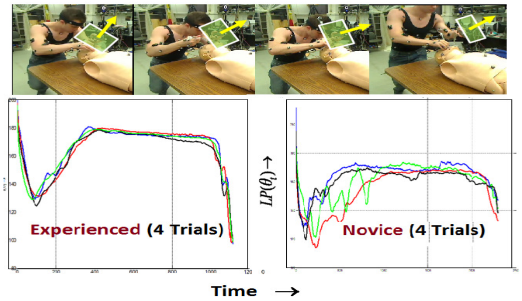


Fig. 5. Inter-trial variability of $LP(\theta_t)$ profile. Multiple trials of a subject were aligned using generalized time warping or GTW [14]

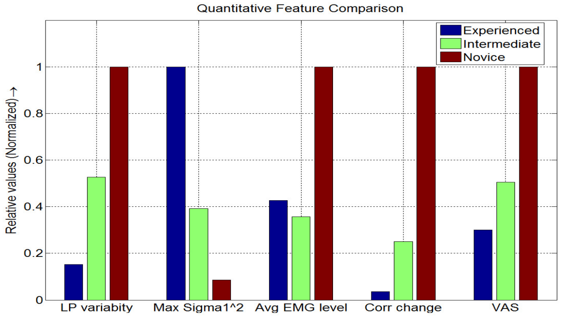


Fig. 6. Relative quantitative comparisons across various features averaged over all trials. LP variability means inter-trial variability of $LP(\theta_t)$ profile. Max σ_1^2 is computed as the peak amplitude of temporal variation of shape deformation parameters. Corr change denoted the number of zero-crossings in the EMG cross-correlation profile. VAS means the number of modes associated with the variance profile of LP angular speed.

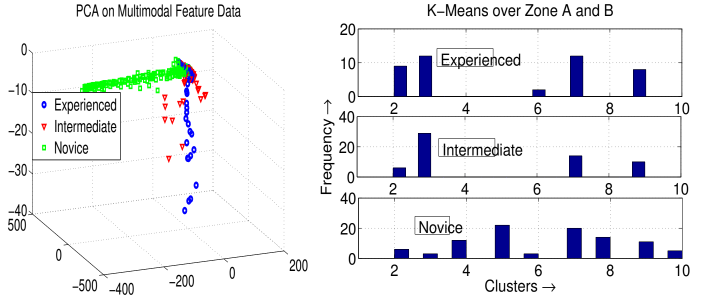
4. REFERENCES

[1] Pepe, P. E., Copass, M. K., and Joyce, T. H.,"Prehospital endotracheal intubation: rationale for training emergency medical personnel, *Ann Emerg Med*, 1985. 14(11): p. 1085-92.

[2] Bushra, J. S et al., "Comparison of trauma intubations managed by anesthesiologists and emergency physicians, *Acad Emerg Med*, 2004. 11(1): p. 66-70.

[3] Hubble, M. W. et al., "A meta-analysis of prehospital airway control techniques part I: orotracheal and nasotracheal intubation success rates, *Prehosp Emerg Care*, 2010.

[4] Wang, H. E. et al.,"Out-of-hospital airway management in the United States, *Resuscitation*, 2011. 82(4): p. 378-85.



a) Visualization after PCA　b) K-means histogram representation

Fig. 7. Fig. a shows 3D visualization of the feature space after PCA. Notice that the data points of the novice is scattered in a direction almost orthogonal to that of the other two subjects. Fig. b shows the histogram representation of ETI zone A and B combined together after k-means clustering. Observe the similarity in the representations of the experienced and the intermediate provider.

[5] O'Connor, R. E. and Megargel, R. E., "The effect of a quality improvement feedback loop on paramedic skills, charting, and behavior, *Prehosp Disaster Med*, 1994. 9(1): p. 35-8;

[6] Carlson, J., et al., "Variables Associated With Successful Intubation Attempts Using Video Laryngoscopy: A Preliminary Report in a Helicopter Emergency Medical Service, *Prehospital Emergency Care*, 2011(under peer review)

[7] Saleh, G. M. et al.,"Kinematic analysis of surgical dexterity in intraocular surgery, *Arch Ophthalmol*, 2009. 127(6): p. 758-62.

[8] McBeth, P. B. et al., "Quantitative methodology of evaluating surgeon performance in laparoscopic surgery, *Stud Health Technol Inform*, 2002. 85: p. 280-6.

[9] I. Dryden and K. Mardia, "Statistical Shape Analysis", *John Wiley and Sons*, 1998.

[10] Das, S. and Vaswani N., "Nonstationary Shape Activities:Dynamic Models for Landmark Shape Changeand Applications, *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol 32. No. 4, April 2010, p. 579-592

[11] Bao, L., and Intille, S. S.,"Activity recognition from userannotated acceleration data , *In Proceceedings of the 2nd International Conference on Pervasive Computing*, 117.

[12] Vicon Motion Systems. Web : "http://www.vicon.com/"

[13] T. Cootes et al. , Active Shape Models-Their Training and Application, *Computer Vision and Image Understanding*, 1995.

[14] Zhou, F. and De la Torre, F., "Generalized Time Warping. (under review)