Deep Eye-CU (DECU): Summarization of Patient Motion in the ICU

Carlos Torres[1(✉)], Jeffrey C. Fried[2(✉)], Kenneth Rose[1], and B.S. Manjunath[1]

[1] University of California Santa Barbara, Santa Barbara, USA
{carlostorres,rose,manj}@ece.ucsb.edu
[2] Santa Barbara Cottage Hospital, Santa Barbara, USA
jfried@sbch.org

Abstract. Healthcare professionals speculate about the effects of poses and pose manipulation in healthcare. Anecdotal observations indicate that patient poses and motion affect recovery. Motion analysis using human observers puts strain on already taxed healthcare workforce requiring staff to record motion. Automated algorithms and systems are unable to monitor patients in hospital environments without disrupting patients or the existing standards of care. This work introduces the DECU framework, which tackles the problem of autonomous unobtrusive monitoring of patient motion in an Intensive Care Unit (ICU). DECU combines multimodal emissions from Hidden Markov Models (HMMs), key frame extraction from multiple sources, and deep features from multimodal multiview data to monitor patient motion. Performance is evaluated in ideal and non-ideal scenarios at two motion resolutions in both a mock-up and a real ICU.

1 Introduction

The recovery rates of patients admitted to the ICU with similar conditions vary vastly and often inexplicably. ICU patients are continuously monitored; however, patient mobility is not currently recorded and may be a major factor in recovery variability. Clinical observations suggest that adequate patient positioning and controlled motion increase patient recovery, while inadequate poses and uncontrolled motion can aggravate wounds and injuries. Healthcare applications of motion analysis include quantification (rate/range) to aid the analysis and prevention of decubitus ulcers (bed sores) and summarization of pose sequences over extended periods of time to evaluate sleep without intrusive equipment.

Objective motion analysis is needed to produce clinical evidence and to quantify the effects of patient positioning and motion on health. This evidence has the potential to become the basis for the development of new medical therapies and the evaluation of existing therapies that leverage patient pose and motion manipulation. The framework introduced in this study enables the automated collection and analysis of patient motion in healthcare environments. The monitoring system and the analysis algorithm are designed, trained, and tested in a mock-up ICU and tested in a real ICU. Figure 1 shows the major elements of the

G. Hua and H. Jégou (Eds.): ECCV 2016 Workshops, Part II, LNCS 9914, pp. 178–194, 2016.
DOI: 10.1007/978-3-319-48881-3_13

framework (stages A–H). Stage A (top right) contains the references. Stage B (bottom left) shows frames from a sample sequence recorded using multimodal (RGB and Depth) multiview (three cameras) sources. At stage C, the framework selects the summarization resolution and activates the key frame identification stage (if needed). Stage D contains the motion thresholds (dense optic-flow estimated at training) to distinguish between the motion types and account for depth sensor noise. Deep features are extracted at stage E. Stage F shows the key frame computation, which compresses motion and encodes motion segments (encoding of duration of poses and transitions). Stage G shows the multimodal multiview Hidden Markov Model trellis under two scene conditions. Finally, stage H shows the results: pose history and pose transition summarizations.

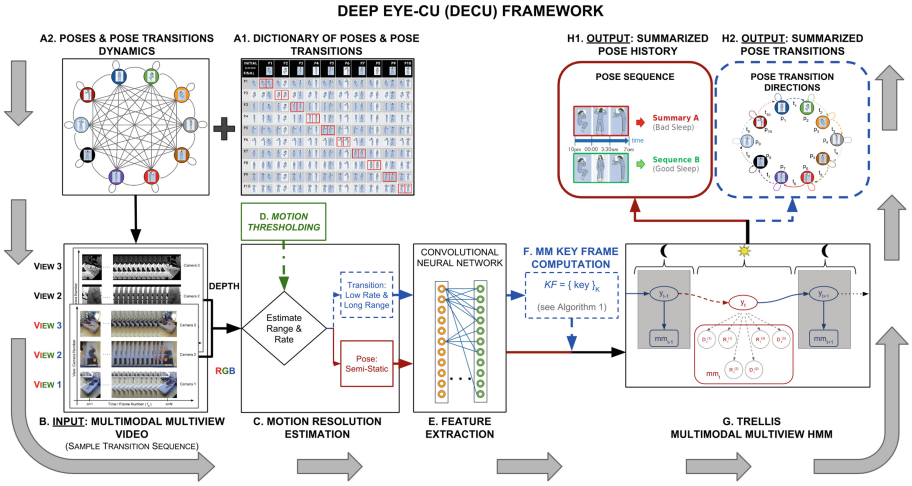


Fig. 1. Diagram explaining the DECU framework, which uses Hidden Markov Modeling and multimodal multiview (MM) data. Stage A provides the references; (A1) a dictionary of poses and pose transitions, and (A2) the illustrative motion dynamics between two poses. Stage B shows the multimodal multiview input video. Stage C selects the summarization resolution and activates key frame identification when required. Stage D integrates the motion thresholds (estimated at training) to account for various levels of motion resolution and sensor noise. Stage F shows the key frame identification process using Algorithm 1. Stage G shows the multimodal multiview HMM trellis, which encodes illumination and occlusion variations. Stage H shows the two possible summarization outputs (H1) pose history and (H2) pose transitions.

Background. Clinical studies covering sleep analysis indicate that sleep hygiene directly impacts healthcare. In addition, quality of sleep and effective patient rest are correlated to shorter hospital stays, increased recovery rates, and decreased mortality rates. Clinical applications that correlate body pose and movement to medical conditions include sleep apnea – where the obstructions of the airway are affected by supine positions [1]. Pregnant women are recommended to sleep

on their sides to improve fetal blood flow [2]. The findings of [3–5] correlate sleep positions with quality of sleep and its various effects on patient health. Decubitus ulcers (bed sores) appear on bony areas of the body and are caused by continuous decubitus positions[1]. Although nefarious, bed sores can be prevented by manipulating patient poses over time. Standards of care require that patients be rotated every two hours. However, this protocol has very low compliance and in the U.S., ICU patients have a probability of developing DUs of up to 80 % [6]. There is little understanding about the set of poses and pose durations that cause or prevent DU incidence. Studies that analyze pose durations, rotation frequency, rotation range, and the duration of weight/pressure off-loading are required, as are the non-obtrusive measuring tools to collect and analyze the relevant data. Additional studies analyze pose manipulation effects on treatment of severe acute respiratory failure such as: ARDS (Adult Respiratory Distress Syndrome), pneumonia, and hemodynamics in patients with various forms of shock. These examples highlight the importance of DECU's autonomous patient monitoring and summarization tasks. They accentuate the need and challenges faced by the framework, which must be capable of adapting to hospital environments and supporting existing infrastructure and standards of care.

Related Work. There is a large body of research that focuses on recognizing and tracking human motion. The latest developments in deep features and convolutional neural network architectures achieve impressive performance; however, these require large amounts of data [7–10]. These methods tackle the recognition of actions performed at the center of the camera plane, except for [11], which uses static cameras to analyze actions. Method [11] allows actions to not be centered on the plane; however, it requires scenes with good illumination and no occlusions. At its current stage of development the DECU framework cannot collect the large number of samples necessary to train a deep network without disrupting the hospital.

Multi-sensor and multi-camera systems and methods have been applied to smart environments [12,13]. The systems require alterations to existing infrastructure making their deployment in a hospital logistically impossible. The methods are not designed to account for illumination variations and occlusions and do not account for non-sequential, subtle motion. Therefore, these systems and methods cannot be used to analyze patient motion in a real ICU where patients have limited or constrained mobility and the scenes have random occlusions and unpredictable levels of illumination.

Healthcare applications of pose monitoring include the detection and classification of sleep poses in controlled environments [14]. Static pose classification in a range of simulated healthcare environments is addressed in [15], where the authors use modality trust and RGB, Depth, and Pressure data. In [16], the authors introduce a coupled-constrained optimization technique that allows them to remove the pressure sensor and increase pose classification performance. However, neither method analyzes poses over time or pose transition dynamics.

[1] Online Medical Dictionary.

A pose detection and tracking system for rehabilitation is proposed in [17]. The system is developed and tested in ideal scenarios and cannot be used to detect constrained motion. In [18] a controlled study focuses on work flow analysis by observing surgeons in a mock-up operating room. A single depth camera and Radio Frequency Identification Devices (RFIDs) are used in [19] to analyze work flows in a Neo-Natal ICU (NICU) environment. These studies focus on staff actions and disregard patient motion. Literature search indicates that the DECU framework is the first of its kind. It studies patient motion in a mock-up and a real ICU environment. DECU's technical innovation is motivated by the shortcomings of previous studies. It observes the environment from multiple views and modalities, integrates temporal information, and accounts for challenging natural scenes and subtle patient movements using principled statistics.

Proposed Approach. DECU is a new framework to monitor patient motion in ICU environments at two motion resolutions. Its elements include time-series analysis algorithms and a multimodal multiview data collection system. The algorithms analyze poses at two motion resolutions (sequence of poses and pose transition directions). The system is capable of collecting and representing poses from multiview multimodal data. The views and modalities are shown in Fig. 2(a) and (b). A sample motion summary is shown in Fig. 2(c). Patients in the ICU are often bed-ridden or immobilized. Overall, their motion can be unpredictable, heavily constrained, slow and subtle, or aided by caretakers. DECU uses key frames to extract motion cues and temporal motion segments to encode pose and transition durations. The set of poses used to train and test the framework are selected from [15]. DECU uses HMMs to model the time-series multimodal multiview information. The emission probabilities encode view and modality information and the changes in scene conditions are encoded as states. The two resolutions address different medical needs. Pose history summarization is the coarser resolution. It provides a pictorial representation of poses over time (i.e., the history). The applications of the pose history include prevention and analysis of decubitus ulcerations (bed sores) and analysis of sleep-pose effects on quality of sleep. The pose transition summarization is the finer resolution. It looks at the pseudo/transition poses that occur while a patient transitions between two clearly defined sleep poses. Physical therapy evaluation is one application of transition summarization. The pose and transition sets are shown in Fig. 1(A1).

Main Contributions

1. An adaptive framework called DECU that can effectively record and analyze patient motion at various motion resolutions. The algorithms and system detect patient behavior/state and healthy normal motion to summarize the sequence of patient sleep poses and motion between two poses.
2. A system that collects multimodal and multiview video data in healthcare environments. The system is non-disruptive and non-obtrusive. It is robust to natural scenes conditions such as variable illumination and partial occlusions.

3. An algorithm that effectively compresses sleep pose transitions using subset of the most informative and most discriminative frames (i.e., key frames). The algorithm incorporates information from all views and modalities.
4. A fusion technique that incorporates the observations from the multiple modalities and views into emission probabilities to leverage complementary information and estimate intermediate poses and pose transitions over time.

2 System Description

The DECU system is modular and adaptive. It is composed of three nodes and each node has three modalities (RGB, Depth, and Mask). At the heart of each node is a Raspberry Pi3 running Linux Ubuntu, which controls a Carmine RGB-D cameras[2]. The units are synchronized using TCP/IP communication. DECU combines information from multiple views and modalities to overcome scene occlusions and illumination changes.

Multiple Modalities (Multimodal). Multimodal studies use complementary modalities to classify static sleep poses in natural ICU scenes with large variations in illumination and occlusions. DECU uses these findings from [15,16] to justify using multiple views and modalities.

Multiple Views (Multiview). The studies from [16,20] show that analyzing actions from multiple views and multiple orientations greatly improves detection and provides algorithmic view and orientation independence.

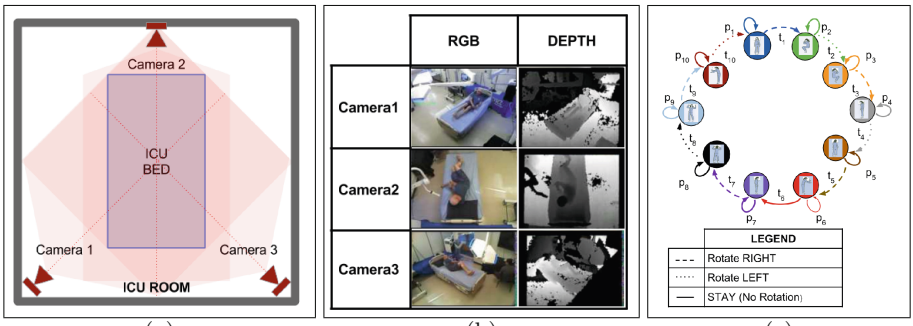
Time Analysis (Hidden Semi-Markov Models). ICU patients are often immobilized or recovering. They move subtly and slowly (very different from the walking or running motion). DECU effectively monitors subtle and abrupt patient motion by breaking the motion cues into temporal segments.

3 Data Collection

Pose data is collected in a mock-up ICU with 10 actors and tested in medical ICU with two real patients (two days worth of data). The diagram in Fig. 2(b) shows the top-view of the rigged mock-up ICU room and the camera views. In the mock-up ICU, actors are asked follow the same test sequence of poses. The sequence is set at random using a random number generator. Figure 2(c) shows a sequence of 20 observations, which include ten poses (p_1 to p_{10}) and ten transitions (t_1 to t_{10}) with random transition direction.

All actors in the mock-up ICU are asked to assume and hold each of the poses while data is being recorded from multiple modalities and views. A total of 28 sessions are recorded: 14 under ideal conditions (BC: bright and clear) and 14 under challenging conditions (DO: dark and occluded).

[2] Primesense, manufacturer of Carmine sensors, was acquired by Apple Inc. in 2013; however, similar devices can be purchased from structure.io.



(a) (b) (c)

Fig. 2. The transition data is collected in a mock-up ICU and a real ICU: (a) shows the relative position of the cameras with respect to the ICU room and ICU bed; (b) shows a set of randomly selected poses and pose transitions, which are represented by lines (dashed, dotted, and solid lines defined in the legend box); (c) shows the complete set of possible sleep-pose pair combinations.

Pose Data. The actors follow the sequence poses and transitions shown in Stage A from Fig. 1. Each initial pose has 10 possible final poses (inclusive) and each final pose can be arrived to by rotating left or right. The combination of pose pairs and transition directions generates a set of 20 sequences for each initial pose. There are 10 possible initial poses. A recording session of one actor generates 200 sequence pairs. Also, two patients sessions are recorded in the medical ICU for one day each (two-hour long video recordings).

Feature Selection. Previous findings indicate that engineered features such as geometric moments (gMOMs) and histograms of oriented gradients (HOG) are suitable for the classification of sleep poses. However, these features are limited in their ability to represent body configurations in dark and occluded scenarios. The latest developments in deep learning and feature extraction led this study to consider deep features extracted from the VGG [21] and the Inception [22] architectures. Experimental results (see Sect. 5) indicate that Inception features perform better than gMOMs, HOG, and VGG features. Parameters for gMOM and HOG extraction are obtained from [15]. Background subtraction and calibration procedures from [23] are applied prior to feature extraction.

4 Problem Description

Temporal patterns caused by sleep-pose transitions are simulated and analyzed using HSMMs as shown in Sects. 4.1 and 4.2. The interaction between the modalities to accurately represent a pose using different sensor measurements are encoded into the emission probabilities. Scene conditions are encoded into the set of states (i.e., the analysis of two scenes doubles the number of poses).

4.1 Hidden Markov Models (HMMs)

HMMs are a generative approach that models the various poses (pose history) and pseudo-poses (pose transitions summarization) as states. The hidden variable or state at time step k (i.e., $t = k$) is y_k (state$_k$ or pose$_k$) and the observable or measurable variables ($x_{k,m}^{(v)}$, the vector of image features extracted from the k-th frame, the m-th modality, and the v-th view) at time $t = k$ is x_k (i.e., $x_k = x_{k,m}^{(v)} = \{R_k, D_k, ...M_k\}$). The first order Markov assumption indicates that at time t, the hidden variable y_t, depends only on the previous hidden variable y_{t-1}. At time t the observable variable x_t depends on the hidden variable y_t. This information is used to compute the joint probability $P(Y, X)$ via:

$$P\big(Y_{1:T}, X_{1:T}\big) = P(y_1) \prod_{t=1}^{T} P\big(x_t|y_t\big) \prod_{t=2}^{T} P\big(y_t|y_{t-1}\big), \tag{1}$$

where $P(y_1)$ is the initial state probability distribution (π). It represents the probability of sequence starting ($t = 1$) at pose$_i$ (state$_i$). $P\big(x_t|y_t\big)$ is the observation or emission probability distribution (**B**) and represents the probability that at time t pose$_i$ (state$_i$) can generate the observable multimodal multiview vector x_t. Finally, $P\big(y_t|y_{t-1}\big)$ is the transition probability distribution (**A**) and represents the probability of going from pose$_i$ to pose$_o$ (state$_i$ to state$_o$). The HMM has parameters $\mathbf{A} = \{a_{ij}\}$, $\mathbf{B} = \{\mu_{in}\}$, and $\pi = \{\pi_i\}$.

Initial State Probability Distribution ($\boldsymbol{\pi}$). The initial pose probabilities are obtained from [4] and adjusted to simulate the two scenes considered in this study. The scene independent initial state probabilities π is shown in Table 1.

State Transition Probability Distribution (A). The transition probabilities are estimated using the transitions from one pose to the next one for Left (L) and Right (R) rotation direction as indicated in the results from Fig. 7.

Emission Probability Distribution (B). The scene information is encoded into the emission probabilities. This information server to model moving from one scene condition to the next shown in Fig. 3. The trellis shows two scenes, which doubles the number of hidden states. The alternating blue and red lines (or solid and dashed lines) indicate transitions from one scene to the next.

One limitation of HMMs is their lack of flexibility to model pose and transition (pseudo-poses) durations. Given an HMM in a known pose or pseudo-pose, the probability that it stays in there for d time slices is: $P_i(d) = (a_{ii})^{d-1}(1 - a_{ii})$, where $P_i(d)$ is the discrete probability density function (PDF) of duration d in pose i and a_{ii} is the self-transition probability of pose i [24].

Table 1. Initial transition probability for each of the 10 poses. Notice that poses facing Up have a higher probability than the poses that face Down, while Left and Right poses are equally probable. Please note that there is a category for poses not covered in this study identifiable by the label Other and the symbol p_{11}. Also, note that one pose can have two states based on the BC and DO scene conditions.

Initial State Probability: $\pi = \{\pi_i\}$						
Pose name	Acronym	Symbol	State - BC	Probability	State - DO	Probability
Soldier up	solU	p1	s_1	0.03	s_{11}	0.02
Fetal right	fetR	p2	s_2	0.145	s_{12}	0.07
Fetal left	fetL	p3	s_3	0.145	s_{13}	0.07
Log right	logR	p4	s_4	0.05	s_{14}	0.03
Soldier down	solD	p5	s_5	0.02	s_{15}	0.01
Yearner left	YeaL	p6	s_6	0.04	s_{16}	0.02
Log left	logL	p7	s_7	0.05	s_{17}	0.03
Faller down	falD	p8	s_8	0.05	s_{18}	0.02
Faller up	falU	p9	s_9	0.05	s_{19}	0.03
Yearner right	yeaR	p10	s_{10}	0.04	s_{20}	0.02
Other	other	p0	s_0	0.036	s_0	0.073

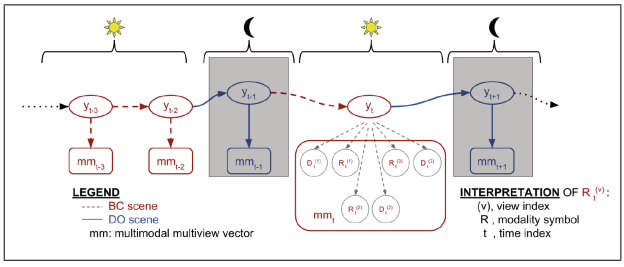


Fig. 3. Multimodal Multiview Hidden Markov Model (mmHMM) trellis. The variation in scene illumination between night and day are examples of scene changes. (Color figure online)

4.2 Hidden Semi-Markov Models (HSMMs)

HSMMs are derived from conventional HMMs to provide state duration flexibility. HSMMs represent hidden variables as segments, which have useful properties. Figure 4 shows the structure of the HSMM and its main components. The sequence of states $y_{1:T}$ is represented by the segments (S). A segment is a sequence of unique, sequentially repeated symbols. The segments contain information to identify when an observation is first detected and its duration based on the number of observed samples. The elements of the j-th segment (S_j) are the indexes (from the original sequence) where the observation (b_j) is detected, the number of sequential observations of the same symbol (d_j), and the state or pose (y_j). For example, the sequence $y_{1:8} = \{1, 1, 1, 2, 2, 1, 2, 2\}$ is

represented by the set of segments $S_{1:U}$ with elements $S_{1:J} = \{S_1, S_2, S_3, S_4\} = \{(1,3,1),\ (4,2,2),\ (6,1,1),\ (7,2,2)\}$. The letter J is the total number of segments and the total number of state changes. The elements of the segment $S_1 = (1,3,1)$ are, from left to right: the index of the start of the segment (from the sequence: $y_{1:8}$); the number of times the state is observed; and the symbol.

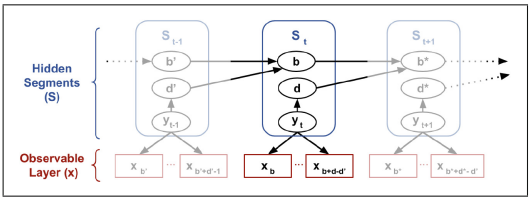


Fig. 4. HSMM diagram indicating the hidden segments S_j indexed by j and their elements $\{b_j, d_j, y_j\}$. The variable b is the first detection in a sequence, y is the hidden layer, (x) is the observable layer containing samples from time b to $b + d - d'$. The variables b and d are the observation's detection (time tick) and duration.

HSMM Elements. The hidden variables are the segments $S_{1:U}$, the observable variables are the features $X_{1:T}$, and the joint probability is given by:

$$P\big(S_{1:U}, X_{1:T}\big) = P\big(Y_{1:U}, b_{1:U}, d_{1:U}, X_{1:T}\big)$$

$$P\big(S_{1:U}, X_{1:T}\big) = P(y_1)P(b_1)P(d_1|y_1) \prod_{t=b_1}^{b_1+d_1+1} P(x_t|y_1)\times$$

$$\prod_{u=2}^{U} P(y_u|y_{u-1})P\big(b_u|b_{u-1}, d_{u-1}\big) \times P\big(d_u|y_u\big) \prod_{t=b_u}^{b_1+d_1+1} P(x_t|y_u), \tag{2}$$

where U is the sequence of segments such that $S_{1:U} = \{S_1, S_2, ..., S_U\}$ for $S_j = \big(b_j, d_j, y_j\big)$ and with b_j as the start position (a bookkeeping variable to track the starting point of a segment), d_j is the duration, and y_j is the hidden state ($\in \{1, ..., Q\}$). The range of time slices starting at b_j and ending at $b_j + d_j$ (exclusively) have state label y_j. All segments have a positive duration and completely cover the time-span $1 : T$ without overlap. Therefore, the constraints $b_1 = 1$, $\sum_{u=1}^{U}$ and $b_{j+1} = b_j + d_j$ hold.

The transition probability $P(y_u|y_{u-1})$, represents the probability of going from one segment to the next via:

$$\mathbf{A} : P\big(y_u = j|y_{t-u} = i\big) \equiv a_{ij} \tag{3}$$

The first segment (b_u) always starts at 1 $(u = 1)$. Consecutive points are calculated deterministically from the previous point via:

$$P\big(b_u = m|b_{u-1} = n, d_{u-1} = l\big) = \delta\big(m, n + l\big) \tag{4}$$

where $\delta(i,j)$ is the Kroenecker delta function (1, for $i=j$ and 0, else). The duration probability is $P(d_u = l|y_u = i) = P_i(l)$, with $P_i(l) = \mathcal{N}(\mu, \sigma)$.

Parameter Learning. Learning is based on maximum likelihood estimation (mle). The training sequence of key frames is fully annotated, including the exact start and end frames for each segment $X_{1:T}, Y_{1:T}$. To find the parameters that maximize $P\big(Y_{1:T}, X_{1:T}|\theta\big)$, one maximes the likelihood parameters of each of the factors in the joint probability. The reader is referred to [25] for more details. In particular, the observation probability $P\big(x^n|y=i\big)$, is a Bernoulli distribution whose max likelihood is estimated via:

$$\mu_{n,i} = \frac{\sum_{t=1}^{T} x_t^i \delta\big(y_t, i\big)}{\sum_{t=1}^{T} \delta\big(y_t, i\big)}, \tag{5}$$

where T is the number of data points, $\delta(i,j)$ is the Kroenecker delta function, and $P\big(y_t = j|y_{t-1} = i\big)$ is the multinomial distribution with:

$$a_{ij} = \frac{\sum_{n=2}^{N} \delta\big(y_n, j\big)\delta\big(y_{n-1}, i\big)}{\sum_{n=2}^{N} \delta\big(y_{t-1}, j\big)} \tag{6}$$

4.3 Key Frame (KF) Selection

Data collected from pose transition is very large and often repetitive, since the motion is relatively slow and subtle. The pre-processing stage incorporates a key frame estimation step that integrates multimodal and multiview data. The algorithm used to select a set (KF) of K-transitory frames is shown in Fig. 5 and detailed in Algorithm 1. The size of the key frame set is determined experimentally ($K = 5$) on the feature scape using Inception vectors.

Let $\mathcal{X} = \{x_{m,n}^{(v)}\}_f$ be the set of training features extracted from V views and M modalities over N frames and let P_i and P_o represent the initial and final poses. The transition frames are indexed by n, $1 \leq n \leq |N|$. The views are indexed by v, $1 \leq v \leq |V|$ and the modalities are indexed by m, $1 \leq m \leq |\mathcal{M}|$. Algorithm 1 uses this information to identify key frames. Experimental evaluation of $|KF|$ is shown in Fig. 5. The idea behind key frames selection is to identify informative and discriminative frames using all views and modalities.

5 Experimental Results and Analysis

Static Pose Analysis - Feature Validation. Static sleep-pose analysis is used to compare the DECU method to previous studies. Couple-Constrained Least-Squares (cc-LS) and DECU are tested on the dataset from [16]. Combining the cc-LS method with deep features extracted from two common network architectures improved classification performance over the HOG and gMOM features in dark and occluded (DO) scenes by an average of eight percent with Inception and four percent with Vgg. Deep features matched the performance of cc-LS (with HOG and gMOM) in a bright and clear scenario as shown in Table 2.

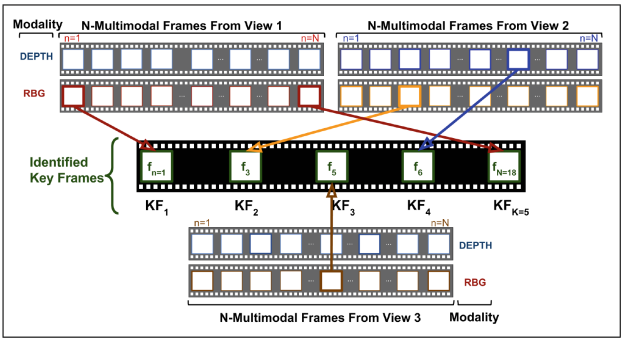


Fig. 5. Selection of transition key frames based on Algorithm 1. This figures shows how the algorithm is used to identify five key frames from three views and two modalities. The first two key frames are extracted from the RGB view 1 video. Subsequent key frames are selected from Depth view 2 and RGB view 3 videos.

Input: $\mathcal{X}$, set of mm features and dissimilarity threshold th;
Result: $KF = \{\text{Key Frames}\}_K$, $K \geq 1$
Initialize: $KF = \{\text{empty}\}_K$, $K \geq 1$ and $count = 0$;
Stage 1: Modality (m) and View (v) Selection;
for $1 < v < V$ *and* $1 < m < M$ **do**
  $D_m^{(v)} = \text{euclid}(x_{mn_i}^{(v)}, x_{mn_o}^{(v)})$, $n_i = 1, n_o = N$;
end
$\hat{v}, \hat{m} = \max D_m^{(v)} > th$;
$\{x_{\hat{m}n_1}^{(\hat{v})}, x_{\hat{m}n_N}^{(\hat{v})}\} \rightarrow FK$;
Stage 2: Find Complementary Frames to KF ;
for $1 < v < V$ *and* $1 < m < M$ *and* $1 < n < N$ **do**
  $D_1 = D_{m,n_1}^{(v)} = \text{euclid}(x_{mn_1}^{(v)}, x_{mn}^{(v)})$;
  $D_2 = D_{m,n_N}^{(v)} = \text{euclid}(x_{mn_N}^{(v)}, x_{mn}^{(v)})$;
end
Sort $D_1 = \{d_1 > d_2 > ... > d_{N-2}\}$ descending;
Sort $D_2 = \{d_1 > d_2 > ... > d_{N-2}\}$ descending;
$d_i \rightarrow KF$ if $\frac{d_i}{d_j} > th$, for $1 < i, j < N - 2$;
Stage 3: Find Center Frame (i.e., Motion Peak);
for KF_2 *and* KF_{K-1} **do**
  Use Stage 2 to compute D_3 and D_4;
  if $max(D_3, D_4) > 0$) **then**
    $\max(D_3, D_4) \rightarrow KF$;
  end
end

Algorithm 1. Multimodal multiview key frame selection using euclidean dissimilarity measure. The algorithm is applied at training with labeled frames to estimate the number and indexes of key frames across views and modalities.

Key Frame Performance. The size of the set of key frames that represent a pose transition affects DECU performance. DECU currently uses $|KF| = 5$ and a dissimilarity threshold $th \geq .8$ as shown in Fig. 6.

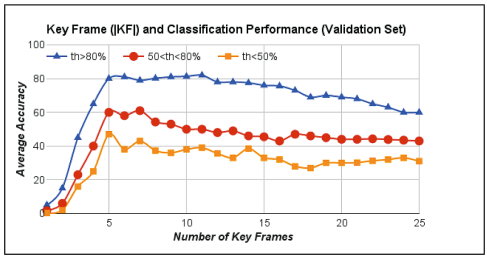


Fig. 6. Performance of the DECU framework for the fine motion summarization based on the number of key frames used to represent transitions and rotations between poses.

Table 2. Evaluation of deep features for sleep-pose recognition tasks using the cc-LS method from [16] in dark and occluded (DO) scenes using. The performance of HOG and gMOM is compared to the performance of the Vgg and Inception features.

Feature Suitability Evaluation with cc-LS [16]			
Scene	HOG + gMOM	Vgg	Inception
BC	100	100	100
DO	65	69 (+4)	73 (+8)

Table 3. Pose history summarization performance (percent accuracy) of the DECU framework in bright and clear (BC) and dark and occluded (DO) scenes. The sequences are composed of 10 poses with durations that range from 10 s to 1 min. The sampling rate is set to once per second.

DECU: Pose History Summarization	
Scene	Average Detection Rate
BC	85
DO	76

Summarization Performance in a Mock-Up ICU Room. The mock-up ICU allows staging the motion and scene condition variations. The sample test sequence is shown in Fig. 2(c).

Pose History Summarization. History summarization requires two parameters: sampling rate and pose duration. The experiments are executed with a sampling rate of one second and an pose duration of 10 s with a minimum average detection of 80 %. A pose is assigned a label if consistently detect 80 % of the time, else they are assigned the label "other". Poses not consistently detected are ignored. The system is tested in the mock-up setting using a randomly selected scene and sequence of poses that can range from two poses to ten poses. The pose durations are also randomly selected with one scene transition (from BC to DO

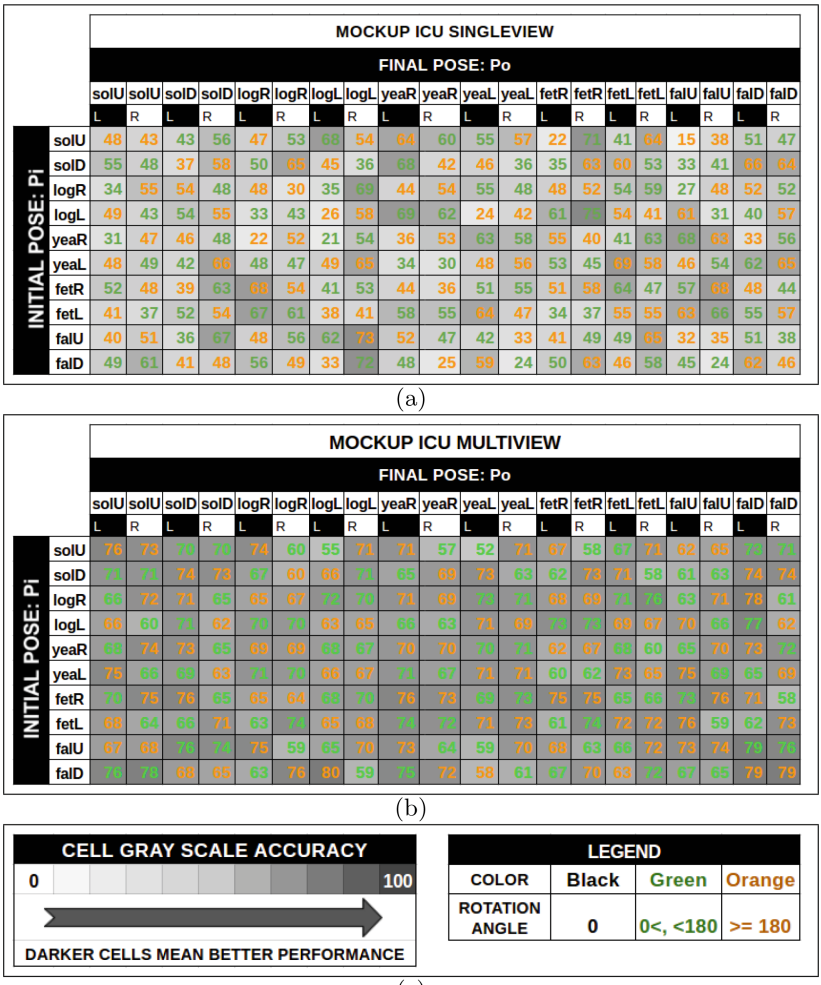
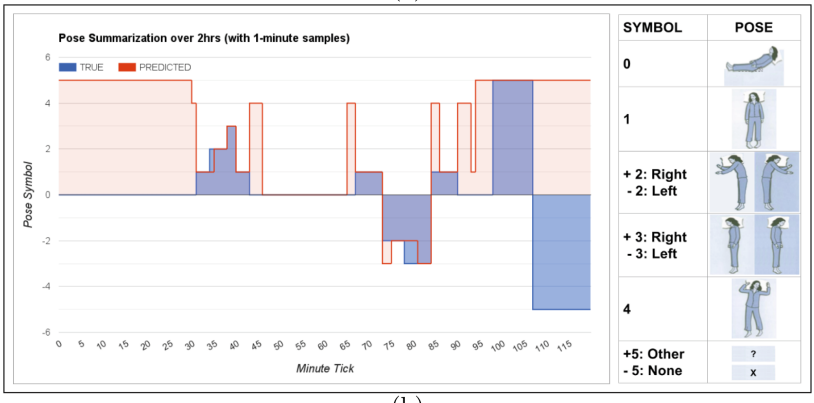


Fig. 7. Performance of DECU in the mock-up ICU under a dark and occluded conditions. Detection results are obtained using (a) single view and (b) multiview data. The cells are gray scaled to indicate detection accuracy. The color coded scale and the legend are shown in (c). Note that overall detection improves with longer rotation angles and worsens when rotations include facing the bed (cameras recording actor backs). (Color figure online)

(a)



(b)

REAL ICU MULTIVIEW

FINAL POSE: Po

INITIAL POSE: Pi

	solU L	solU R	logR L	logR R	logL L	logL R	yeaR L	yeaR R	yeaL L	yeaL R	fetR L	fetR R	fetL L	fetL R	falU L	falU R
solU	N/A		73	88	N/A		70	65	N/A		71	80	N/A		57	
logR	79	N/A			80	N/A	63		N/A	84	62		84	N/A	76	N/A
logL	N/A	73	N/A	83	N/A		76		68		N/A	86	61		N/A	79
yeaR	81	N/A	60		81	N/A			83	N/A	58		81	N/A	78	N/A
yeaL	N/A	79	N/A	83	60		N/A	80	N/A			75	64		N/A	82
fetR	83	N/A	58		81	N/A	67		82	N/A			78	N/A	86	N/A
fetL	N/A	77	N/A	87	69		N/A	85	63		N/A	87	N/A		72	
falU	60		N/A	72	78	N/A		77	72	N/A		76	79	N/A		

(c)

Fig. 8. Performance of DECU pose transition summarization in a real ICU shown in (a) using multimodal data under natural scene conditions. The set of patient poses is reduced and the summarization performance for a two hour session is shown in (b). The detection scores are shown in (c), where the cells are gray scaled to indicate detection accuracy. The font color indicates rotation angle range and N/A indicates the pose is not available (i.e., not possible). The grading color scale is shown in Fig. 7(c). (Color figure online)

or from DO to BC). A sample (long) sequence is shown in Fig. 2(c) and its history summarization performance is shown in Table 3.

Pose Transition Dynamics: Motion Direction. The analysis and pose transitions and rotation directions are important to physical therapy and recovery rate analysis. The performance of DECU summarizing fine motion to describe transitions between poses is shown in Fig. 7. Results for the DO scene with (a) singleview and (b) multiview data. The legend is shown in (c).

Summarization Performance in a Real ICU. The medical ICU environment is shown in Fig. 8(a) and (b). Note that it is logistically impossible to control ICU work flows and to account for unpredictable patient motion. For example, ICU patients are not free to rotate, which reduces the set of pose transitions (unavailable transitions are marked N/A). The set of poses for the history summary require that a new pose be included (pulmonary aspiration). A qualitative illustration is shown in Fig. 8(b). DECU's fine motion summarization results for two patients are shown in Fig. 8(c).

6 Conclusion and Future Work

This work introduced the DECU framework to analyze patient poses in natural healthcare environments at two motion resolutions. Extensive experiments and evaluation of the framework indicate that the detection and quantification of pose dynamics is possible. The DECU system and monitoring algorithms are currently being tested in real ICU environments. The performance results presented in this study support its potential applications and benefits to healthcare analytics. The system is non-disruptive and non-intrusive. It is robust to variations in illumination, view, orientation, and partial occlusions. DECU is non-obtrusive and non-intrusive but not without a cost. The cost is noticed in the most challenging scenario where a blanket and poor illumination block sensor measurements. The performance of DECU to monitor pose transitions in dark and occluded environments is far from perfect; however, most medical applications that analyze motion transitions, such as physical therapy sessions, are carried under less severe conditions.

Future studies will investigate the recognition and analysis of patient motion and interactions in natural hospital scenarios using recurrent neural networks and integrate natural language understating to log ICU actions and events.

Acknowledgements. This research is sponsored in part by the Army Research Laboratory under Cooperative Agreement Number W911NF-09-2-0053 (the ARL Network Science CTA). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on. The authors thank

Dr. Richard Beswick (Director of Research), Paula Gallucci (Medical ICU Nurse Manager), Mark Mullenary (Director Biomedical-Engineering), and Dr. Leilani Price (IRB Administration) from Santa Barbara Cottage Hospital for their support.

References

1. Sahlin, C., Franklin, K.A., Stenlund, H., Lindberg, E.: Sleep in women: normal values for sleep stages and position and the effect of age, obesity, sleep apnea, smoking, alcohol and hypertension. Sleep Med. **10**, 1025–1030 (2009)
2. Morong, S., Hermsen, B., de Vries, N.: Sleep position and pregnancy. In: de Vries, N., Ravesloot, M., van Maanen, J.P. (eds.) Positional Therapy in Obstructive Sleep Apnea. Springer, Heidelberg (2015)
3. Bihari, S., McEvoy, R.D., Matheson, E., Kim, S., Woodman, R.J., Bersten, A.D.: Factors affecting sleep quality of patients in intensive care unit. J. Clin. Sleep Med. **8**(3), 301–307 (2012)
4. Idzikowski, C.: Sleep position gives personality clue. BBC News, 16 September 2003
5. Weinhouse, G.L., Schwab, R.J.: Sleep in the critically ill patient. Sleep-New York Then Westchester **10**(1), 6–15 (2006)
6. Soban, L., Hempel, S., Ewing, B., Miles, J.N., Rubenstein, L.V.: Preventing pressure ulcers in hospitals. Joint Comm. J. Qual. Patient Saf. **37**(6), 245–252 (2011)
7. Chéron, G., Laptev, I., Schmid, C.: P-cnn: pose-based cnn features for action recognition. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 3218–3226 (2015)
8. Veeriah, V., Zhuang, N., Qi, G.J.: Differential recurrent neural networks for action recognition. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 4041–4049 (2015)
9. Baccouche, M., Mamalet, F., Wolf, C., Garcia, C., Baskurt, A.: Sequential deep learning for human action recognition. In: Salah, A.A., Lepri, B. (eds.) HBU 2011. LNCS, vol. 7065, pp. 29–39. Springer, Heidelberg (2011)
10. Tran, D., Bourdev, L., Fergus, R., Torresani, L., Paluri, M.: Learning spatiotemporal features with 3d convolutional networks. In: 2015 IEEE International Conference on Computer Vision (ICCV), pp. 4489–4497. IEEE (2015)
11. Soran, B., Farhadi, A., Shapiro, L.: Generating notifications for missing actions: don't forget to turn the lights off! In: Proceedings of the IEEE International Conference on Computer Vision, pp. 4669–4677 (2015)
12. Hoque, E., Stankovic, J.: Aalo: activity recognition in smart homes using active learning in the presence of overlapped activities. In: 2012 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops, pp. 139–146. IEEE (2012)
13. Wu, C., Khalili, A.H., Aghajan, H.: Multiview activity recognition in smart homes with spatio-temporal features. In: Proceedings of the Fourth ACM/IEEE International Conference on Distributed Smart Cameras, pp. 142–149. ACM (2010)
14. Huang, W., Wai, A.A.P., Foo, S.F., Biswas, J., Hsia, C.C., Liou, K.: Multimodal sleeping posture classification. In: IEEE International Conference on Pattern Recognition (ICPR) (2010)
15. Torres, C., Hammond, S.D., Fried, J.C., Manjunath, B.S.: Multimodal pose recognition in an icu using multimodal data and environmental feedback. In: International Conference on Computer Vision Systems (ICVS). Springer (2015)

16. Torres, C., Fragoso, V., Hammond, S.D., Fried, J.C., Manjunath, B.S.: Eye-cu: sleep pose classification for healthcare using multimodal multiview data. In: Winter Conference on Applications of Computer Vision (WACV). IEEE (2016)
17. Obdržálek, S., Kurillo, G., Han, J., Abresch, T., Bajcsy, R., et al.: Real-time human pose detection and tracking for tele-rehabilitation in virtual reality. Stud. Health Technol. Inform. **173**, 320–324 (2012)
18. Padoy, N., Mateus, D., Weinland, D., Berger, M.O., Navab, N.: Workflow monitoring based on 3d motion features. In: 2009 IEEE 12th International Conference on Computer Vision Workshops (ICCV Workshops), pp. 585–592. IEEE (2009)
19. Lea, C., Facker, J., Hager, G., Taylor, R., Saria, S.: 3d sensing algorithms towards building an intelligent intensive care unit, vol. 2013, p. 136. American Medical Informatics Association (2013)
20. Ramagiri, S., Kavi, R., Kulathumani, V.: Real-time multi-view human action recognition using a wireless camera network. In: ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC) (2011)
21. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
22. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9 (2015)
23. Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision, 2nd edn. Cambridge University Press, New York (2004)
24. Rabiner, L.R.: A tutorial on hidden markov models and selected applications in speech recognition. Proc. IEEE **77**(2), 257–286 (1989)
25. Van Kasteren, T., Englebienne, G., Kröse, B.J.: Activity recognition using semi-markov models on real world smart home datasets. J. Ambient Intell. Smart Env. **2**(3), 311–325 (2010)