2015 IEEE International Conference on Services Computing

CloudMat: Context-aware Personalization of Fitness Content

(Invited Paper)

Kuk Jin Jang
Electrical & Systems Engineering
University of Pennsylvania
jangkj@seas.upenn.edu

Jungmin Ryoo
Dept. of Bioengineering
University of Pennsylvania
ryooj@seas.upenn.edu

Orkan Telhan
School of Design
University of Pennsylvania
otelhan@seas.upenn.edu

Rahul Mangharam
Electrical & Systems Engineering
University of Pennsylvania
rahulm@seas.upenn.edu

Abstract*—Digital video content via broadcast television, Internet and other content distribution networks provide limited interaction for fitness and wellness activities. The content delivery is one-way only and provides no personalization to the pace, programming and progress of the user's exercise routine. Furthermore, the content is to be viewed only on a screen which makes it awkward and incompatible with full-body activities such as yoga, pilates and T'ai chi. We present *CloudMat*, a system for context-aware personalization of fitness content with cloud-enabled connected surfaces. CloudMat provides real-time closed-loop feedback between the state of the user on the physical mat and the state of the content in the cloud service. Content is tagged with actuation signals where events are delegated from the screen to display on an electroluminescent lighting layer on the mat, which provides spatial guidance to the end-user. Through the sensor-layer embedded in the mat, the physical interface captures the pose and timing of the user activity and relays it to the Context-aware Personalization cloud service. This service coordinates sensing and actuation between the content stream and mat by generating pose templates and metadata files about the exercise routine to be delivered to the user. Through this interactive process between the physical mat and the content service, the feedback provided by the user performing the routine continuously adapts the pace and programming to maintain the desired user experience. We demonstrate the utility of the system and evaluate the system performance with a case study on interactive yoga.

I. INTRODUCTION

More content is consumed by digital media than any other form of media today. Digital media provides superior quality in audio and video encoding, efficient delivery modes, on-line mixing of multiple streams, and improved analytics. However, digital video content delivery still provides only limited forms of interaction that restrict it to passive screen-based consumption, which is largely one-way communication. Previous attempts of interactive TV failed [1], [2] because they focused on convergence of multiple forms of communication and challenged the passive nature of video content consumption. Convergence is fundamentally a flawed concept, but it's even worse when you try to put an active medium (e.g. the Internet) together with a passive medium (e.g. television.)

When video content is used for fitness and wellness activities such as yoga, pilates or t'ai chi, the *screen-based* interface presents severe limitations on the interaction as the user is often on all four limbs and is unable to easily view the screen. Screen-based viewing assumes a sitting or standing posture which is incompatible with a full-body fitness routine where

The first two authors contributed equally

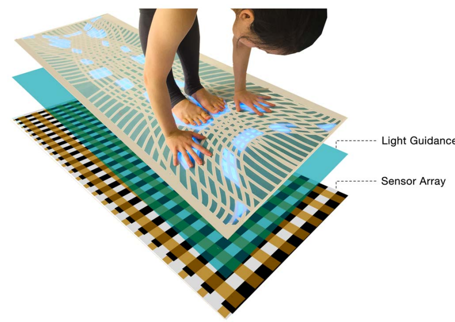


Fig. 1. Physical interface of CloudMat. The structure of the mat consists of a sensor array layer, actuation layer with lighting guidance and top surface

the person's face may be away from the screen. A second issue is that the *pace* of broadcast content is fixed and does not adapt to the user's needs. The delivery is *one-way* as there is no mechanism for interaction. Ideally, the pace at which the content is delivered to the interface should adjust to the pace of the user as she progresses through the fitness routine. Interaction between the content and the user will require a means for feedback from the user's environment. Finally, the content is *"one size fits all"* in that the modules within a show's episode are the same for all users. As the physical capabilities are different across the population of persons participating in fitness activities, we would ideally want content that changes its pace and sequencing based on feedback from the user's environment. For example, if a very fit person's workout routine involves weights, then she may focus just on upper body exercises for 35 minutes. However, a novice may start to fatigue with upper body exercises after a few minutes. The sensors in the weights will detect fatigue and the content will adapt by sequencing abdomen exercises followed by lower body exercises for a shorter duration. This way, the sensors in the user's environment provide feedback so the content's pace and modules may match the user's capabilities.

The focus of this effort is on mechanisms and infrastructure for context-aware personalization for digital media content delivery. Personalization means that the content adapts to the user's environment and vice versa by adjusting the pace of content delivery by responding to interactive sensor-actuator interfaces that are alternatives to the screen and matches content sequencing to the user's needs and preferences. While new means of inter action are introduced both on the side of the physical world with new interfaces and the content-side, there is no requirement for specializing or changing the content in any way. In other words, the cost for producing the content remains the same for these value added services.

978-1-4673-7281-7/15 $31.00 © 2015 IEEE
DOI 10.1109/SCC.2015.49

Unlike traditional DOCSIS 3.0 delivery of digital media content which is largely one-way (Fig. 2 (left)), context-aware personalization offers personalized experiences specific to the current audience and the media type (Fig. 2 (right)). This is achieved by closing the loop by inferring the state in the user's physical environment through sensors and adapting the content within the Content Cloud Service. The state of the content is also reflected in the physical environment through actuators which signal the user for a response. The control signaling between the sensors, Content Cloud Service and actuators is achieved by synchronized out-of-band meta-data channels which exchange the states of the physical and content worlds. This approach allows for new and alternative interaction interfaces that are no longer limited to the visual and audio components of the content. By targeting individual preferences and expanding the modes of delivery it is our aim that the user will naturally be more engaged in the content, resulting in a more effective presentation and satisfying viewing experience.

We focus on context personalization for full-body fitness and wellness activities such as yoga, pilates and T'ai chi. For example, in 2012, the yoga market grew to $7 billion and spending on yoga-related is predicted to exceed $8 billion by 2017 [3]. It is popular by with beginners due to its low skill and strength barrier for entry, is easy to do at home and in a studio alike and is also recognized in incentive-based corporate wellness programs. Despite the popularity, the content available for yoga is limited to pre-recorded sequences and do not feature any interactive guidance.

In this work, we present CloudMat, a *content-coupled* interactive yoga and wellness mat which senses user activity, infers movements based on yoga contexts, interacts with a cloud service to facilitate the generation context-aware fitness content and allows for personalized experiences based on real-time feedback to the user through actuators within the mat. The following are contributions of this paper:

1) Interactive Interfaces: A novel, cloud-enabled connected surface made of piezo-resistive material as a seamless sensing modality and electroluminescent lighting for guidance during exercise.

2) Context Inferencing: An automated pose template generation to train the inference engine with yoga sequences and a pose recognition algorithm to extract statistically relevant of user activity.

3) Closed-loop Cloud Service: A closed-loop system enabling personalized fitness experiences with real-time feedback for pacing, content sequencing and performance analytics.

With such closed-loop interaction we aim to extend legacy one-way static video content and screen-based delivery to an interactive experience on responsive interfaces for fitness and wellness activities.

In the following sections, we begin with an overview of the concept of context-aware personalization and content-coupled interaction for fitness content. The system overview and implementation details are explained in Section III and IV. In Sections V and VI, we evaluate the inferencing engine and discuss performance analytics across a population of users.

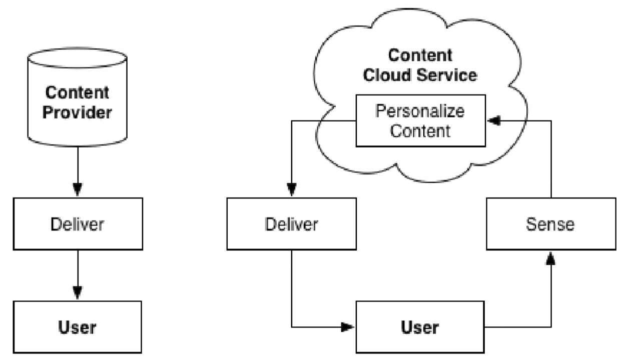


Fig. 2. Structure of Traditional Digital Media Delivery (left) and Context-aware Personalized Content Delivery (right)

II. Context-Aware Personalization for Fitness and Wellness

A. Three Types of Content-coupled Personalization

Ironically, the content delivery infrastructure for cable-TV and Internet video do not contain significant actionable information on the context or contents of the media. The context of digital media consists of information about the content itself (e.g. what is happening in the scene) and the environment where it is currently being consumed (e.g. physical room occupied by the user). The goal of this effort is to develop mechanisms for conveying this context and physical interfaces that record and respond to such contexts. For fitness-related digital media, the context includes the actions performed by the user during a type of exercise as well as elements from the environment such as the current skill level of the user and the content module queue order in response to the real-time feedback provided by the user during exercise. Context-aware fitness personalization can be summarized in the following key aspects of personalization:

1) Temporal Personalization: Temporal personalization allows dynamic adjustment of pacing during playback of fitness media. Segments of the content can be paused or skipped according to the real-time feedback provided by the user's environment. For example, in a yoga video, the content can be paused automatically to wait for a user to achieve a step in a routine. If the inferencing engine detects user fatigue via the sensors on the CloudMat, the content can adapt appropriately, for example, by skipping certain modules and/or selecting modules with less intensive exercises. This allows a more natural and personalized interaction during the routine.

2) Spatial Personalization: Spatial personalization is achieved by extending the delivery of fitness content from the screen to the physical environment of the user through physical interfaces such as the interactive lighting embedded in the fitness mat presented here. By projecting the video content as the active limb and body positions on the mat, the user is given guidance and feedback in the physical realm. The mapping of content to illuminations on the mat is calibrated to the physical constraints of the user (e.g. height and flexibility captured during the calibration sequence). This is analogous to a (simplified) real-life physical trainer who adjusts the specific actions and poses of an exercise routine to suit the user.

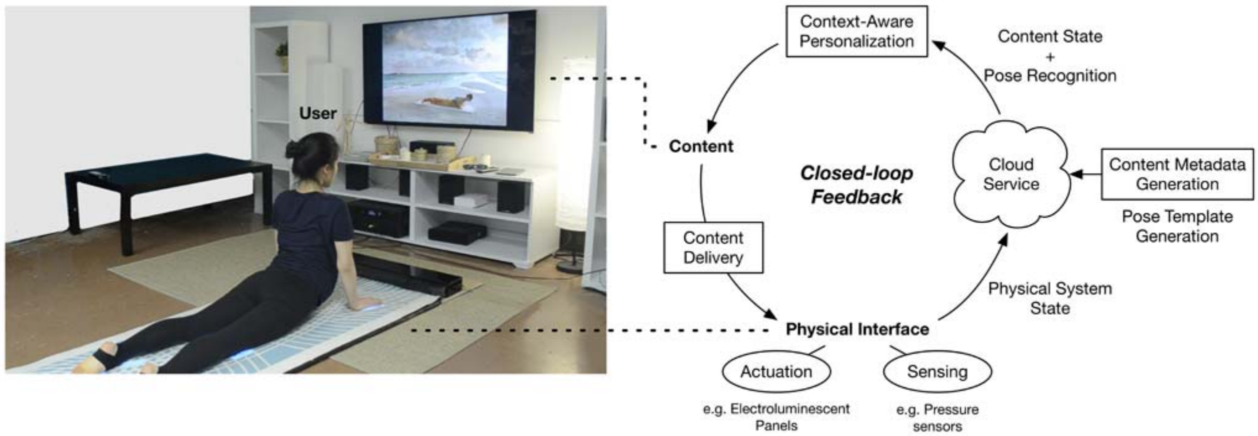


Fig. 3. CloudMat System Architecture. The system consists of 3 stages: Content Generation, Content Delivery, and Context-aware Personalization

3) Procedural Personalization: Procedural personalization encourages active engagement with the content over a longer period of time. Based on metrics evaluating the current fitness level of the user, the target routine may be adjusted to increase the difficulty and continuously garner interest in the user.

III. CloudMat Overview

A. Content-coupled Internet of Things

To couple the content with physical systems, such as a yoga mat, we have developed a Coupled-content Cloud Service which exchanges metadata regarding a description of the current scene in the content stream to the physical system. In the case of a yoga program, the metadata encodes the sequence of movements within the current pose which enumerates limb and body locations on the mat and also encodes pull requests from the sensors on the mat to learn the user's current position. This metadata labeling of the content is done manually by the domain expert once and does not affect the legacy content stream. The metadata is delivered as an out-of-band data stream that is synchronized with the content delivery.

The yoga mat itself has an actuation layer with electroluminescent lighting which lights up according to the positions in the metadata stream. The mat also has a sensing layer (as seen in Fig. 1 which captures the user's current position. The inference engine connected to the mat determines the user's progress based on the sequence of poses recognized for the current routine.

Given this closed-loop operation between the sensors on the yoga mat, the Coupled-content Cloud Service infers the user's progress and personalizes the content for (a) user-specific pacing, (b) mapping of context to alternate interfaces (e.g. electroluminescent lights on the mat) and (c) sorting the queue's content modules to suit the user's preferences.

B. Content-coupled Yoga Mat System Architecture

As shown in Fig. 3, system operation is divided into three stages of content metadata generation, content delivery and context-aware personalization. The system includes an interactive fitness mat which was developed as the physical interface for the system. The mat consists of a sensor array layer to receive feedback from the user and a lighting guidance layer to deliver personalized content in the form of spatial personalization. The system collects and processes physical data from the user during exercise to aid both in the metadata generation for context-aware fitness content delivery and provides real-time feedback to the user.

C. Context-aware Fitness Content Personalization Process Overview

Context-aware fitness content personalization follows the three stages of the system. First, a fitness instructor records content while using the interactive mat. For a given range of video frames, the relative positioning and placement of the limbs and body in the form of pressure heat maps is captured in the metadata stream. Using this data, the Pose Template Generation engine (described in the next section) creates pose templates which are annotated with body profile information of the instructor. The content is delivered to the user in the form of a JSON metadata stream which actuate the lighting guidance provided by the interactive mat. Feedback obtained from the user through the mat's sensors is processed with knowledge of the current pose templates. Descriptions of the hardware and software components of the system are detailed in subsequent sections.

D. Use Case: Interactive Yoga

In this paper, the system was applied to yoga as a proof-of-concept demonstrating the utility of the system and verifying the functionality of key components. Yoga was chosen as the representative exercise to demonstrate the system as it is relatively structured with sets of pose sequences, where each pose has a well-defined placement of body and pacing between different poses. This allows us to verify the operation of the content-coupled yoga mat.

IV. System Implementation

This section will describe each of the stages which comprise the system in detail. Fig. 1 shows the structure of the physical interface for CloudMat. The physical interface consists of 3 layers: a top layer with cover design, a lighting guidance layer based on printed electroluminescent ink, and a sensor array layer. In addition to the physical interface, the

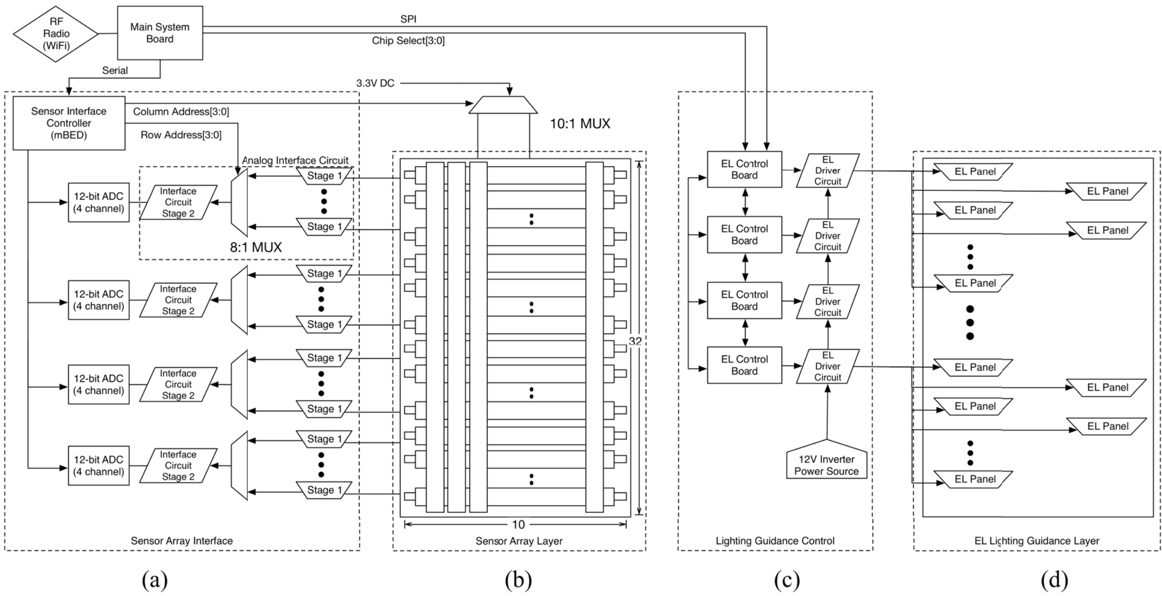


Fig. 4. CloudMat Block Diagram. The physical interface and control system can be divided the main system board and the following subsystems and layers: (a) sensor array interface, (b) sensor array layer (c) lighting guidance layer and (d) lighting guidance control

system includes a graphical interface for data acquisition, as well as the software components for template generation and pose recognition.

A. Cloud-enabled Connected Surface

1) Sensor Array Layer: The sensor array layer of the interactive mat is a modular sensor array of dimension ($24in. \times 72in.$). As shown in Fig. 4(b), the sensor array has a structure similar to [4]. A piezoresistive polymer sheet of Velostat with a thickness of 101.6μm and a volume resistivity of $500\Omega/cm$ is inserted between the two adhesive external sheets, where the conductive rows and columns are made of copper foil. The Velostat has been cut in stripes to inhibit the intra-row eddy currents induced by the sensor array interface circuit. Each crossing of a row and column acts as a sensor element. The spacing and the width of the rows and the columns is determined to be the maximum values that distinguish the left and right, and front and back side of the average size of the adult hand and foot over a $2in. \times 2in.$ and $3in. \times 2in.$. The resistance of the sensor element area changes with the application of pressure.

2) CloudMat Control System: The control system board is depicted in Fig. 4. It consists of the sensor array interface board, the lighting guidance control, and the main system board which communicates with the host system.

a) Sensor Array Interface: As shown in Fig. 4(a), the sensor array interface divides the mat into 4 sections of 8 rows with each row connected to an analog voltage divider interface circuit which transduces the variable resistance into a voltage signal. Each row of a section is time-multiplexed into a second stage analog interface circuit which performs conditioning and filtering. During the time for a row, an input voltage of 3.3V is time-multiplexed across the 10 columns of the sensory array. The signals from the 4 sections are fed into 4 analog input channels of the mBED microcontroller board (NXP LPC1768,100 MHz ARM Cortex-M3) and sampled at a rate of 200kHz at 12-bit resolution. The entire mat is sampled every 100ms, which is a reasonable rate for monitoring yoga as the pace compared to other exercises is slower and the change is body placement is less. The microcontroller communicates with the main system board via serial interface.

b) Lighting Guidance Control: The lighting guidance control system is shown in Fig. 4(c). 32 channels are divided across 4 electroluminescent (EL) sequencer boards (ATMega 328p @8MHz) and each channel controls an individual EL panel. Each of the boards are connected in a master-slave configuration through an SPI interface to communicate with the main system board at a speed of 1Mbits/s. When a command for a pose is received from the host, a set of channels are set high and the AC voltage input to the system drives those EL panels turning it on. The panels are connected to the individual output channels of the sequencer board (Fig. 4(d)). This lighting layer provides spatial guidance as to where to place limbs and body for a particular pose in a sequence.

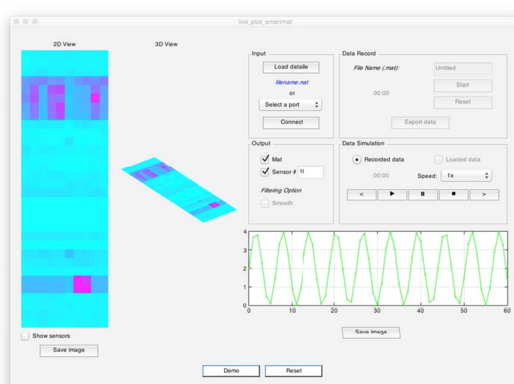
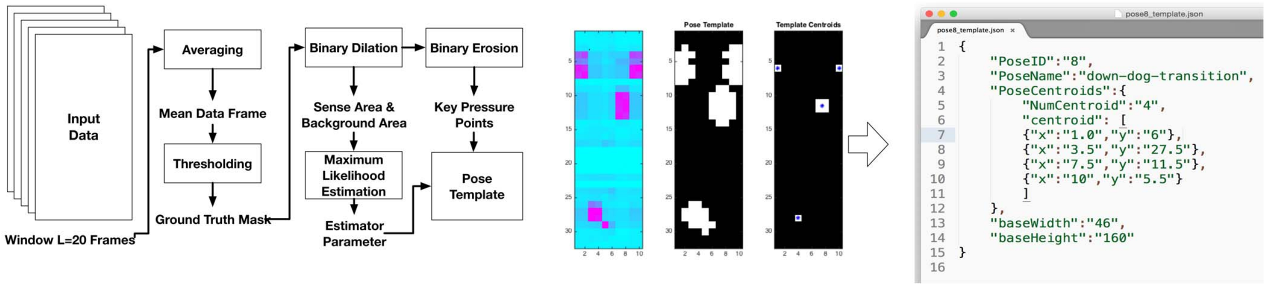


Fig. 5. Mat Sensor Array dashboard. The interface allow for recording of user pose data and generation of pose templates to be used for pose recognition



(a) (b) (c)

Fig. 6. Pose Template Generation Process (a) Block diagram of process for creating a pose template from the data recorded for pose. (b) Example of pose template generated for pose 8 in the sequence (c) Resulting output of the pose template generation automatically converted and stored in JSON format to be utilized by the system during content delivery.

c) Main System Board: The main system board is based on the Spark Core (STM32F103CB ARM 32-bit Cortex-M3, 72Mhz, TI CC3000 Wi-fi module) and controls the two subsystems related to the sensor array and lighting array. The main system receives commmands from the host system to actuate poses on the mat and sends data received from the sensor array interface to the host system. The main system board has built-in Wi-Fi capability that allows for it to communicate wirelessly via a cloud service.

B. Data Acquisition Graphical Interface

In order to test and verify the functioning of the system a graphical user interface (GUI) was implemented in MATLAB to collect the data from the mat and visualize the input in real time. Fig. 5 shows a screen shot of the interface.

C. Offline Pose Template Generation

A key goal of the system is to facilitate generation of pose templates of mat activity. These templates are used during content delivery as a reference for the system to process the sensor input from the user and provide appropriate feedback. Each template represents a key snapshot of activity which the user must complete. Configuring the templates by ordering them with timestamps allows a specific yoga routine to be generated. This template routine is tagged in the metadata stream with the appropriate digital media content timestamp and polls the mat at the the timestamp to process the feedback. Fig. 6(a) depicts the processing of input data from the sensor array. Data is collected in frames of data which are recorded.

Mask Generation. For a particular pose, a window of length L=20 frames of data are averaged together to create *mean data frame*. This frame is thresholded to create a *ground truth mask*, where the set of data points in the mean data frame which are greater than the threshold form the *sense area* and those less than the threshold form the *background area*. The sense area is the region of the mean data frame where the desired pressure actuation should occur by the user for the given pose and the background area is the a region where no pressure should be sensed during the pose. The threshold is determined experimentally and is adjusted according to the pressure distribution created by the person generating the template. The ground truth data mask is processed as a logical array where the 1's are stored in the locations of the data points of the sense area and zero otherwise. Fig. 7 shows the template generation algorithm.

Binary image dilation is performed in the 4-connected to include a greater number of point of the sense area to compensate for the hard threshold A maximum-likelihood estimation algorithm is used to fit a gaussian distribution to the set of data points in the sense are and the background area, respectively. These parameters and the ratio of the sense area to the total frame area and the ratio of the background frame area used as the respective prior probabilities are used to classify the data points in data frame from a user during content delivery.

Maximum-likelihood Estimation. After the estimation of the maximum-likelihood parameters, a binary image erosion is performed to reduce the size of the sense area regions. The centroids of the resulting regions are defined as the key pressure points encoding the template. Finally the template data is converted into a JSON format and stored. The template is used during pose recognition to determine if the user of the system has successfully attained a pose of an exercise routine. An example of the result of processing for pose 8 is shown in Fig. 6(b) and (c).

The pose template generation method of the system greatly simplifies the complexity needed to encode an exercise routine to be done during content-delivery to the end user. Some additional annotations to the template, such as designation of hands and feet on the template, can be recorded to be used as additional criteria for the pose recognition stage.

D. Runtime Pose Recognition

Pose recognition is based on the template generated according to the method described in the previous section.

Classification and Centroid Calculation. A window of length $L = 5$ data frames is combined and intensities averaged to for a *test data frame*. Using the calculated parameters from the template and the prior probabilities, a maximum-likelihood classifier is used to assign data points in the test data frame to either the sense area of the background area. Next, based on the classification, the binary image of the test data frame is derived and connected data points within the sense area are group together as connected regions. Erosion in performed on binary image to remove extraneous regions and as with the template generation, the centroids of the regions are calculated.

Template Matching. The centroids are used as the key pressure points of the test data frame. Afterwards, a point correspondence is performed between the key points of template and the key points of the test data frame. Finally, the distances

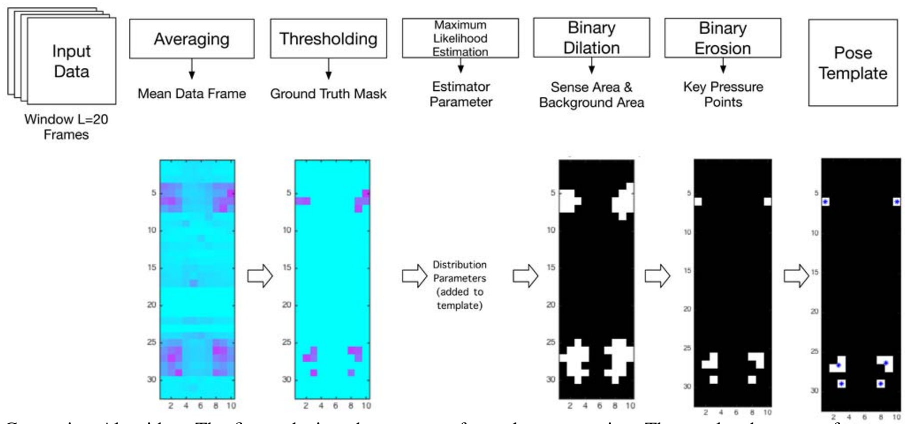


Fig. 7. Pose Template Generation Algorithm. The figure depicts the process of template generation. The results shown are for pose 7.

between the corresponding points are calculated and summed as the mean square error between the template and the test data frame. A thresholding method is used to determine if the test data frame matches the template for the given pose. The threshold determined by the body profile of the user and the body profile of the instructor who generated the content is used to determine if the test data frame matches the template.

During point correspondence, certain assumptions are made about the environment and the user of the system. The pose recognition will target the typical intended use of the system and that the user will be motivated to follow the given instructions for a pose. This means that the pressure source of the input regions for a data frame recorded from the mat will be similar to the pressure source for when the template was generated. The hands, feet and other areas of the body will be placed in locations relative to each other according the pose that is instructed at a given time in the exercise sequence. This assumption is intended to simplify the point correspondence stage and will later be relaxed in future work.

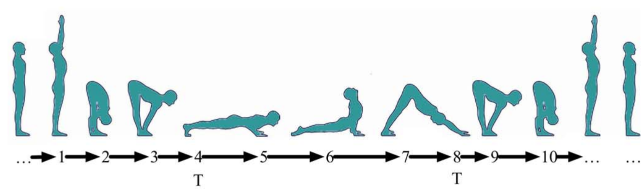
Spatial Personalization and Temporal Personalization. The pose recognition stage utilizes input about the current user during content-delivery which enables for spatial personalization of the required pose. By utilizing this key pressure point-based template matching method, the input data given by a user who has a different body profile from that instructor who generated the template for the pose can still be recognized. Temporal personalization is based on the ratio between success or failure of the user to follow sequences of poses. This ratio is used as an input to the overall system to adjust the pacing of the content. If the ratio of successes falls below a threshold based on user level, the content delivery will slow or pause to match the pace of the user.

V. SYSTEM EVALUATION & EXPERIMENTS

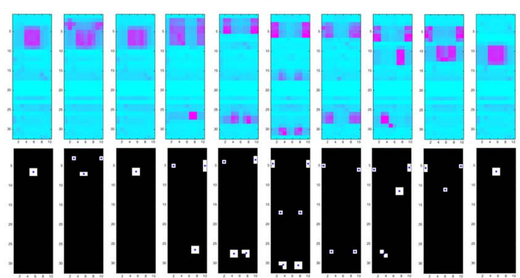
A. Template Generation Parameter Estimation

For pose template generation, a certified ashtanga yoga instructor (co-author Ryoo) was asked to perform 10 predefined yoga sequence called the “sun salutation” to use as the base content. Fig. 8(a) shows the sequence of poses used for the experiment. In the figure, poses 4 and poses 8 are marked with a ’**T**’ representing transitional, intermediary steps in the sequence before the main poses of 5 and 8, respectively. The instructor was asked to create two sequences, one more suited for *beginner-level* users and one targeting *expert-level* users. The difference between routines was mainly in the requirements for selected poses which have variations based on difficulty. For example, pose 6, the *upward-dog* pose, the front surface of the legs and lower torso is not supposed to touch the ground, but for beginners some support by the lower body is allowable.

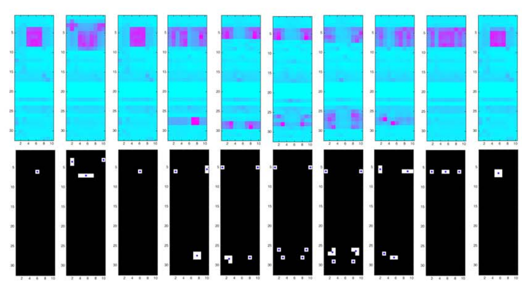
Each pose in a sequence was recorded for twenty seconds. Each recording was divided into intervals of 1 second for



(a)



(b)



(c)

Fig. 8. (a) Diagram of ‘sun salutation’ yoga pose sequence used for system evaluation. Poses marked with ‘T’ are intermediary poses required during transition to a main pose in the sequence that are not depicted explicitly. (b) Template generation results for the beginner-level sequence. (c) Template generation results for the expert-level sequence.

which a mean data frame was extracted to represent the interval. Cross-validation was executed on the data frame set for each pose in the sequence to search for the optimal threshold parameter for template generation and pose recogntion. The parameters which resulted in the best differentiation between difficulty in key poses of a sequence were selected. The results of cross-validation are shown in Table I.

B. System Pose Recognition

Using the generated templates for each routine 3 healthy test subjects (age 24-30) with no prior yoga experience were asked to follow the 10-pose sequence on the sensor mat. Each pose, including intermediary steps were maintained for 20 seconds. Mean data frames were extracted from intervals of 1 second from each recording for a pose. Each subject was asked to repeat the entire sequence two times, once for the *beginner-level* sequence and the *expert-level* sequence. The results for the *expert-level* sequence is detailed in Table II.

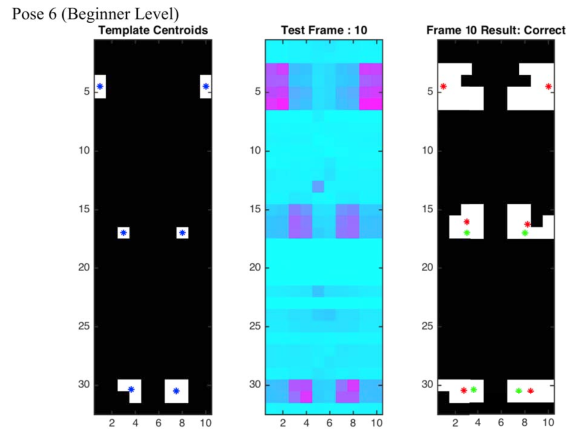
The primary evaluation metric for determining whether a subject had successfully performed a pose in the sequence was the alignment and placement of the feet and hands with respect to the physical dimensions of the subject. During an execution of the sun salutation sequence it is known that the alignment of feet and hands, keeping the distance between two hands shoulder width, achieving balance between supporting hands, and maintaining a stable position during 5 breath counts is an important criteria for evaluating the quality of execution. The chosen evaluation metric was determined to be the most basic form of evaluation and was most suitable for verifying the functionality of the system. The created sensor data is shown in Fig 8.

VI. RESULTS & DISCUSSION

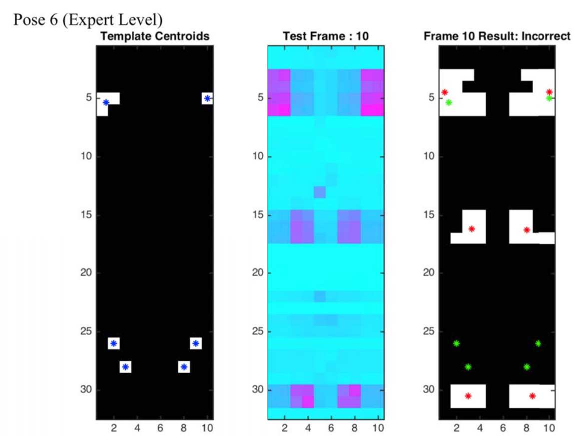
A. Pose Template Generation

Fig. 8 (b) and (c) show the results of template generation for the beginner-level and expert-level sequences. Fig. 8 (b) shows the templates generated for the beginner-level sequence and Fig. 8 (c) shows the templates generated for the expert-level sequence. The templates were generated focusing on the differentiation between the beginner-level and expert-level sequence poses. For many of the more simple poses, such as poses 1 to 3 or poses 8-9, little difference can be made from between the two difficulties based on the location of the pressure point centroids for the template. This is reasonable since the variation in sequence does not necessarily occur for every pose in the sequence. However, other parameters, such as the threshold for error during recognition can be used to adjust the conformance required by the user. For example, from the cross-validation results for Pose 8 in Table I, despite the similarity in the pose templates, for a beginner-level template, the requirements are less strict and so during recognition of a beginner-level sequence, data frames from Pose 8 of the expert-level routine are classified as correct.

For the expert-level sequence, however, the frames from the beginner-level routine are not classified as correct. For other specific poses, such as Pose 6, the difference in the pose requirements based on the difficulty are more apparent. The results for cross-validation are summarized in Table I. As can be seen in the average results, the parameter settings allow



(a) Pose 6 - Beginner Level Routine (Correct)



(b) Pose 6 - Expert Level Routine (Incorrect)

Fig. 9. Pose Recognition Results - Pose 6 - The system differentiates pose recognition based on the routine difficulty. For the beginner level routine (a) the system classifies the pose as correct, while for the expert-level routine (b) the system classifies the pose as incorrect.

for emphasis in recognition of data frames from the respective difficulties. The requirements were more strict for the expert-level templates and routine as can be seen in the results. The data frames from the beginner-level sequence which were classified as correct for the expert-level routine generally came from poses which are similar regardless of difficulty.

B. Pose Recognition

In general, the system was able to recognize the poses and evaluate a given pose as correct or incorrect when the subject successfully maintained the target pose. The system was also able to differentiate based on difficulty level of the routine. For example, in Fig. 9 for Pose 6, for the same pose generated by the beginner-level user, the system classifies the pose as correct during the beginner routine (9a) and incorrect during the expert routine (9b). For pose recognition based on the subject tests only the results for the expert-level routine are reported in Table II. Most of the difference between subjects came from

TABLE I. CROSS-VALIDATION RESULTS FOR TEMPLATE GENERATION AND POSE RECOGNITION PARAMETERS

	Beginner Template Recognition Rate										
Pose	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**AVG**
Beginner Routine	1.0	0.94	1.0	1.0	1.0	0.95	1.0	1.0	1.0	1.0	0.99
Expert Routine	1.0	0	1.0	0.57	1.0	1.0	0.55	0.48	0	1.0	0.66
	Expert Template Recognition Rate										
Pose	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**AVG**
Beginner Routine	1.0	0	1.0	1.0	0.94	0	0	0	0	1.0	0.49
Expert Routine	1.0	1.0	1.0	0.90	1.0	0.95	0.9	1.0	0.91	1.0	0.97

the evaluation based on the expert-level routine. This indicates that the statistics based on the recognition rate could be utilized to measure the skill level of a user. The results output from the system could be utilized for procedural personalization and based on the success rate of the user for a sequence, different poses of varying difficulty could be combined together to form routines which match the user's current skill level.

C. Discussion

The goal for evaluation in this paper was to demonstrate its utility or generation of content and metadata for context-aware fitness personalization. For this paper, the evaluation of pose recognition was centered around the placement and alignment of the pose. The recognition algorithm was a template matching algorithm with an emphasis on simplicity such that implementation in low-cost hardware would be more feasible. Some assumptions were made about the recognition process. First, it was assumed that a user of the system would be actively attempting to participate in the exercise and so at a given time during the routine, the user would be maintain the appropriate pose. This simplifies the algorithm such that the system needs to determine if the user is in one specific pose or not, rather than having to classify the pose out of a set of possible poses the user could be in. Also, the ratio of the physiological dimensions of the test subject to the instructor and the resulting dimensions for the pose were simplified to a linear model. This model allowed for reasonable performance of pose recognition.

A more extensive study of the physical dimensions of the user and the relation with the 2D projection of a user in a particular pose and developing a better inference model would help the robustness of the algorithm. Additionally, the recognition algorithm could be extended to account for other factors such as balance, pressure distribution, etc. in order to create a more robust algorithm and also evaluate other aspects of user performance. If the system were incorporated into a cloud-computing service, such higher-order processing could be delegated to the back-end, and allow for a more extensive evaluation of the user.

TABLE II. PERFORMANCE BY SUBJECT

	Test Subject		
Pose	**A**	**B**	**C**
1	1.0	1.0	1.0
2	0.14	0.0	0.79
3	1.0	1.0	1.0
4	1.0	1.0	1.0
5	0.0	0.88	0.0
6	0.0	0.0	0.0
7	1.0	0.3	1.0
8	0.58	0.0	0.0
9	0.83	0.0	0.0
10	1.0	1.0	1.0
AVG	0.66	0.52	0.58

Apart from these limitations, the entire workflow of generating templates to testing with users was simple with little need for manual adjustment of settings. Moreover, the mat-based interface was intuitive and easy to use for both the test subjects and the yoga instructor. The process is streamlined and modular that it could be possible for the entire system to be incorporated into a cloud-service realizing context-aware content-delivery with personalized feedback. The real-time feedback can be provided based of measurements and user evaluation results from the system. In addition, the mat-based interface is generic and can be extended to other types of exercises, particularly exercise routines utilizing flat surfaces, such as squats and push-ups. These directions for development of the system will be left for future work.

VII. CONCLUSION

In this paper, we presented CloudMat for context-aware fitness personalization in the use-case of interactive yoga. CloudMat provides an intuitive cloud-enabled connected surface interface and simplifies the generation of metadata necessary for context-aware fitness personalization. In addition, the system is able to provide lighting-based spatial guidance to the end user and evaluate the performance of the user based on the real-time feedback of the user based on the generated metadata.

ACKNOWLEDGMENT

The authors would like to thank Yash Pant and Srinivasan Ekembaram at the University of Pennsylvania for their insight during the design of the pose recognition algorithm. In addition, Matthew O'Kelly at the University of Pennsylvania assisted in the construction of physical interface.

REFERENCES

[1] J. Carey, "Winky Dink to Stargazer: Five Decades of Interactive Television," in *I-TV*, 1996.

[2] J. F. Jensen, "Interactive television - a brief media history," in *Proc. of the 6th European Conf. on Changing Television Environments*, ser. EUROITV '08. Springer-Verlag, 2008, pp. 1–10.

[3] A. McElvoy, "Yoga stretches up," Nov. 2014. [Online]. Available: http://www.economist.com/news/21631963-plentiful-pranayama-yoga-stretches-up

[4] M. Sundholm, J. Cheng, B. Zhou, A. Sethi, and P. Lukowicz, "Smart-mat: Recognizing and Counting Gym Exercises with Low-cost Resistive Pressure Sensing Matrix," in *Proc. of the 2014 ACM Intl. Joint Conf. on Pervasive and Ubiquitous Comp.*, ser. UbiComp '14. ACM, 2014, pp. 373–382.