*XXVI Simpozijum o novim tehnologijama u poštanskom i telekomunikacionom saobraćaju – **PosTel 2008**, Beograd, 16. i 17. decembar 2008.*

NON-INVASIVE DATA ACQUISITION FOR USER MODELING

Andrej Košir, Tkalčič Marko, Jurij Tasič
Faculty of Electrical Engineering,
University of Ljubljana, Slovenia

Abstract: *One of the most impending drawbacks of modern information technology and communication devices is the problem of user interfacing. The focus of the research community to tackle this problem is almost exclusively set to the concept of personalization based on user modeling procedures. The major difficulty of user modeling procedures is the required input data about the modeled user. Studies and practical experiences show that, on one hand, the efficiency of user modeling relies on acquired data about the user, and on the other hand, it is extremely irritating for user to constantly provide such data about her or his feelings and behavior. Therefore there is a need for non-intrusive and non-irritating data collection about the user. There are several approaches such us distributed sensors and real time analysis of video capturing the user's behavior with the intention of identifying specific information about user's behavior.*

The problem statement of noninvasive data acquisition for user modeling along with a typical system architecture will be presented in the paper. A set of unavoidable information processing techniques in this context will be introduced. The proposed approach will be further detailed by data acquisition based on a real time video analysis and higher order post processing of gathered user data. A major part of the paper is focused on post processing of user data in terms of geometrical, topological and logical analysis with the aim of achieving an effective user model.

Keywords: *User modeling, noninvasive data acquisition, data fusion*

1. Introduction

The goal of user modeling is to make the interaction between the user and a complicated communication device as easy as possible. Clearly, the user who has bought a device for entertainment, education or any other purpose wants to use it for that purpose and not to read several pages of complicated instructions and then face the even more complicated user interface. To achieve this, we track user interactions over time, acquire and store different kind of information on user properties, features and behavior, both short and long term, and build his user model. We update the user model automatically. Any contextual information available is also used.

Unfortunately, the process of gathering user information is usually irritating for the user and this is in contradiction of the goal of user modeling. Therefore, there is a need for non-invasive (non-intrusive) acquisition of user modeling. For already listed reasons, there are several limitations of such acquisition since we do not wish to bother the user and we wish to respect his privacy. Beside that, the measurement conditions (environment illumination etc) may vary significantly. As a consequence, there is also a need for gathered data postprocessing in order to enhance the performance of the user modeling system. As detailed later, such data processing is the incorporation of contextual apriori available information, data fusion of information from different sensors, etc.

2. Problem statement

The problem of user modeling is to support and ease the interaction between the user and the communication device utilizing the apriori known conceptual information and the analysis of gathered data about the user past behavior. As already mentioned, the problem we address in this paper is how to gather user information without disturbing his work in any way. As such, we continue with the distinction of intrusive vs. Nonintrusive data acquisition and then provide the framework of user modeling system in order to put user modeling into the real world context.

2.1 Intrusive vs. non-intrusive data acquisition

When we speak about non-intrusive (non-invasive) data acquisition we have to define first the criteria for intrusiveness. As our work is focused on human-computer interaction we formulate the following statement:

An approach for data acquisition is considered intrusive when it distracts the user from her/his core interaction with the computer for the sake of data acquisition thus making the core interaction non-spontaneous.

Let us take a movie recommender system as the application with which the user interacts. Asking the user for explicit ratings of a movie is an intrusive form of feedback acquisition since the user is annoyed and distracted from the watching of the movie. A camera that observes the user is considered a non-intrusive way of data acquisition as the user is mostly unaware of it. A body sensor (e.g. skin conductance response sensor) that is physically attached to the user is considered an non-intrusive feedback acquirement channel.

2.2 The framework for user modeling

To put the problem user modeling into the context, we depict the underlying environment including user and data acquisition devices, see Figure 1.

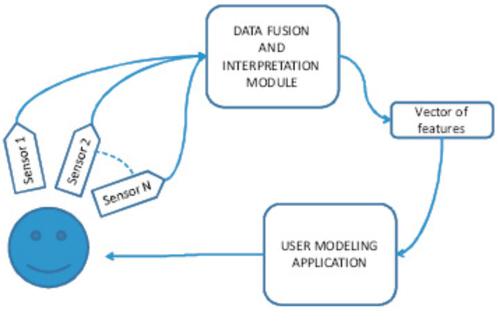
DATA FUSION AND INTERPRETATION MODULE
Vector of features
USER MODELING APPLICATION
Sensor 1
Sensor 2
Sensor N

Figure 1: The topmost view of the proposed framework for adaptive Human-Computer Interaction (HCI). At the basis is the interaction between the user and the application running on a device (computer, set-top-box . . .). The user's response is monitored through a variety of sensors (video stream, Galvanic skin Response, . . .). The sensors' outputs are merged into the *Data fusion and Interpretation module*. This module processes all the inputs from the sensors and produces a vector containing high-level features. It applies machine learning algorithms to perform the requested task. The application (with which the user interacts) is a user-state-aware application and adapts its processing and output to the state of the user. The HCI loop is thus closed.

3 . User scenarios

To clarify the goal of user modeling, we briefly describe one of the many possible user scenarios being recommender system for multimedia items. Other user scenarios are creation of virtual communities, personal shopping assistant, user adopted learning, personalized turist guide etc.

3.1 Scenario: Recommender system for multimedia items

Today we face an impressive growth of digital multimedia content (DVDs, DivXs, MP3s . . .), devices for the consumption of multimedia items (digital TVs, DVD players, set-top-boxes, mobile phones, MP3 players, . . .) and distribution channels (cable TV, optical fiber, WiFi, UMTS, . . .). This puts the end user in a historically unique position: never has she/he been in such a favorable position in terms of choice variety but still, never she/he has been so confused regarding her/his choices.

From the content provider point of view the users are not consuming as much multimedia content as one would think at first glance. This is why recommender systems that are able to filter out from a huge multimedia database only those items that are relevant for the specific end user are starting to get their fair share.

A recommender system is usually formalized as a rating prediction module. Based on knowledge about the end user the recommender system makes a prediction of the rating that the user would give to each item if she/he was able to watch all items [1]. It then makes a limited selection of few top prediction-rated items and offers them to the end user thus easing the choice and boosting consumption. In order to perform the rating predication a recommender system requires knowledge about the end users and knowledge about the content items. These are usually formalized as *user profiles* and *item profiles*. Based on this knowledge the recommender system engages an algorithm, which can be content-based, collaborative or a hybrid combination of both.

An ideal to which we are trying to get is to have accurate feedback gathered in a non-intrusive fashion. To achieve this we propose a platform that collects data from users while they are consuming multimedia content through a wide range of sensors: video camera, GSR (Galvanic Skin Response) sensors, heart pulse, blood pressure, haptic and many others. Figure 2 depicts the proposed solution. The raw inputs are preprocessed and low level features are extracted. Based on machine learning algorithms these low level features are mapped to an emotive state which is a peace of information that the recommender system uses to properly model the user. So far the research in our group is dealing only with video input for the detection of the emotive state and with MPEG7 [2] usage history information. In the proposed architecture we also have explicit user ratings which are used to create a training dataset for tuning the machine learning algorithms.

The detection of the affective state is performed into the VAD (valence-arousaldominance) space. The VAD space is a generalisation of the basic six emotions [3] proposed by Ekman [4].

4. Visual data acquisition and postprocessing

In this section, we briefly present relevant data acquisition techniques and relevant real time processing techniques and then devote more effort into postprocessing techniques being an important part of this work.

4.1 Real time video acquisition

Existing real time video processing techniques are applied in user monitoring and user model building. The quality of USB web cameras now meet the requirements of at least some real time video processing techniques in this context. It is important to bear in mind that the goal of user modeling systems is to help user interacting with complicated communication devices and not to complicate it. Therefore it is important that we do not relay on user collaboration with the video acquisition device since it is usually irritating for him and in contradiction with the purpose of the whole system.

4.2 Real time feature postprocessing

Real time feature postprocessing is of key importance of the proposed system. This is due to the following two facts, 1. the data acquisition conditions varies significantly and may be extremely poor (for instance, the illumination of the scene when video is captured) and 2. the success rate of all known pattern recognition techniques (identity recognition etc) are highly dependent on the quality of the captured data. On the other hand, various types of context dependent information is available and can be utilized to improve the overall performance of the system. For demonstration purposes only we will mostly limit ourselves on real time video based human identity and age recognition for the purpose of user modeling.

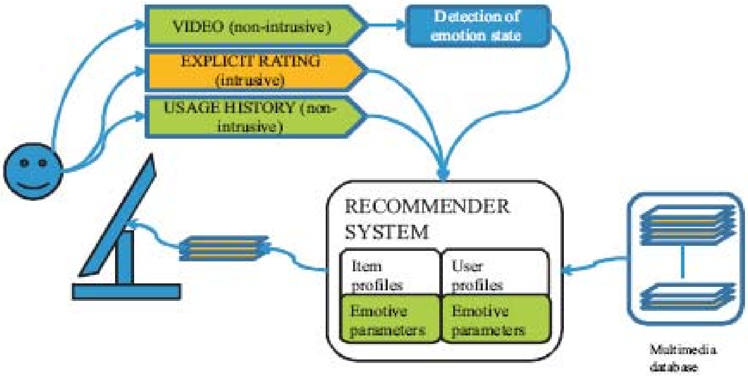

Figure 2: Scenario for the multimedia recommender system: The end user is trying to access multimedia items (movies, pictures) from a huge database through an application. The application implements a recommender algorithm which filters through only those items that are relevant to the user. This is done through a content-based or collaborative content filtering algorithms which are based on the personal user's profile and on items' profiles. The user's response during the consumption of the multimedia item is monitored through 3 channels: the usage history channel, the explicit rating of the item and through a non-intrusive video channel. The video channel is decomposed into low-level features and an emotional state vector is calculated. The information from all three channels is then merged and used by the recommender system to update the profiles for more accurate future content filtering.

In practice, more than one user is monitored simultaneously. In the case of a video on demand application, typically more than one user is watching the selected movie at the same time. Such users can be of different characteristics (age, personal preferences etc) and separate user models are to be built for each user.

A detailed description of user data postprocessing dataflow will be given. Prior to it, let us consider what kind of apriori known information is available and applicable in this context. We divide it into three types as follows.

- Geometrical information: The geometry of the user space (his environment where he consumes the material, e.g. watch movies etc) is known to some extent.
- Topological information: As a distinction from geometrical one, topological information is independent from the monitored scene geometry. It includes impossible configurations like one face below another one and not closer (smaller) enough etc.
- Logical information: An example of logical information is that a particular person can not appear twice at the same frame.

The above described information can be utilized to enhance the quality of selected feature recognition (identity, age, etc) since we can apriori eliminate candidates for human faces according to a contextual map (see geometrical information above). It turned out, details are given in section Discussion later, that the precision of recognition can be increased from useless to good quality one. Beside that, the computational complexity of the whole process can be reduced since the search space is reduced according to the contextual map. Furthermore, the logical postprocessing prevents the system to make obviously wrong recognitions which is extremely corruptive for the usability of built user model.

The architecture of user data postprocessing in the form of a reduced UML diagram is shown at Figure 3. To reduce the size of the image, dataflow activities are not shown but only objects, that is activity results are listed. To provide a detailed explanation, activities (in this case, algorithms and procedures) are labeled by lowercase roman numbers. The dataflow concentrates on users identity based on a real time video analysis only.

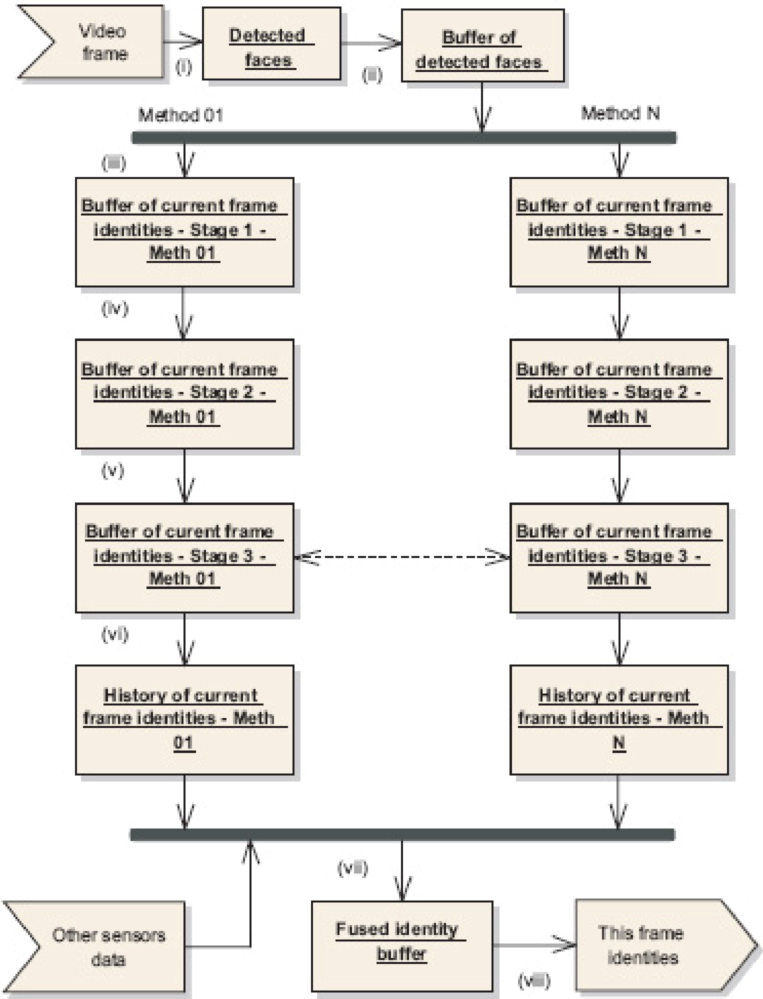
Video frame
(i)
Detected faces
(ii)
Buffer of detected faces
Method 01
Method N
(iii)
Buffer of current frame identities - Stage 1 - Meth 01
Buffer of current frame identities - Stage 1 - Meth N
(iv)
Buffer of current frame identities - Stage 2 - Meth 01
Buffer of current frame identities - Stage 2 - Meth N
(v)
Buffer of curent frame identities - Stage 3 - Meth 01
Buffer of current frame identities - Stage 3 - Meth N
(vi)
History of current frame identities - Meth 01
History of current frame identities - Meth N
(vii)
Other sensors data
Fused identity buffer
(viii)
This frame identities

Figure 3: User data processing dataflow

After some initial remarks we describe the given dataflow according to the mentioned roman number labels of underlying procedures. Note that more than one human face identity recognition method is used in the presented scheme. The left hand side column represents Method 1 and the right hand side column represents Method N assuming that there are N methods available.

First of all, why we use more than one recognition method instead of a single effective one? It turns out that any (identity, age, etc) recognition method has certain drawbacks no matter how well it is tuned up. On the other hand, it turns out that these drawbacks are such that they come out at different situations. Therefore, it is reasonable to run several recognition methods in parallel and then combine their results on each detected human face.

We continue the presented dataflow (Figure 3) description by separate steps of processing labeled by roman numbers.

(i) The goal of this step is to detect all human faces that appear at the input frame. The result is reported as a list of coordinates of rectangles (regions) of human faces. They are of different positions and can be of different sizes. As the most effective general purpose human face detector we use Viola-Jones detector [5], the implementation provided in OpenCV library [7].

(ii) First all detected faces that are not possible in terms of geometrical and topological analysis are removed. Positions and dimensions of all detected faces (regions as rectangles) are then compared to the stored ones and decided to be stored as new ones or already detected ones. If the position and sizes of newly detected face is close enough to one of the stored one, it is assumed that the same face is represented by the new one and the position and dimensions of the previously stored face is set to the new one. Otherwise, if no match is found between in-thisframe detected face and already stored ones, it is stored as a new one. Therefore it is assumed that a new person entered the monitored scene.

(iii) From the step (iii) to the step (vii) face images are processed separately by *N* recognitions methods, certain postprocessing routines are interconnected as depicted by dotted arrows. At this stage, each recognition method is run on each detected face (subregion of captured frame) and a predefined number of most probable identities are stored in the buffer at Stage 1. It turns out that since the recognition is not perfect, it is a good strategy to store not only the best (most probable) identity but a given number of best identities.

(iv) Logical analysis of captured identities is performed to get to the stage 2. First, the multiple occurrences of a single identity are removed in a way such that the most probable ones are chosen. A Bayes probability scheme [24] is applied to determine the most probable ones. The probability of each identity is also stored to the buffer for later use.

(v) At the next stage, identities of the same face detected at the current frame computed by recognition methods 1 to *N* at stage 2 are compared to each other for all detected faces. It may happen that they do not coincide. In this case, the most probable identity of each face is determined in the context of all methods. Again, a Bayes probability scheme is applied here.

(vi) All person (face) identities together with their probabilities determined in previous steps are now added into a buffer holding a history of identities of each face detected on the monitored scene. That is, every detected is tracked over several frames, their identities are determined in each frame and stored into this buffer. This is done for every method separately.

(vii) At step (vii), postprocessed results of all N methods are fused together into the final decision of identity of each identified face on a monitored scene. Statistical data analysis methods are applied here. Histograms of identities for each detected face by each recognition method are built and then statistical decision theory is applied to decide the most likely identity as the final result of the whole procedure. Note that data captured by other sensors and relevant for a given person identity can be taken into account here.

(viii) Final decision is reported into the system to utilize it in building each user model.

5. Discussion

The first question we need to address in discussion is what the gain of the presented postprocessing techniques is. For demonstration purposes only, assume we use PCA [8] based method of human face recognition as a part of automatic user modeling construction. It is known that this method achieves quite a reasonable precision when human faces are frontal and that it's precision goes down rapidly when the analyzed face is rotated. Clearly, there are a lot of rotations present when a typical user behavior is monitored. Experimental results show that with no postprocessing the recognition results are not useful at all. The implementation of the above given scheme provides close to perfect results. In numbers, only 15% of correct identifications can be improved to over 80% of final precision. A general experience shared by many biometrical and surveillance applications is that one can achieve better results by combining several less effective fast recognition methods than by one sophisticated and more effective recognition method.

6. Conclusion and further work

A problem of effective user modeling based on a non-intrusive user data acquisition and gathered data postprocessing is presented in the paper. A real world context of user modeling is introduced in order to motivate the described scheme. According to our experience, the incorporation of contextual information about the user behavior and his working environment is of key importance here. As a consequence, this work mostly concentrates on the postprocessing of gathered user data via real time video analysis. The postprocessing is based on an effective combination of several methods run in parallel and statistical analysis of obtained results.

Our future work plan includes effective data fusion of yet another type of data made available by new sensor technology such as skin moisture measurement. An important pursue is also covering different user scenarios and applications of user modeling that are expected to appear in the area of modern communications.

References

[1] Adomavicius, G. and Tuzhilin, A.Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions, IEEE Transactions on Knowledge and Data Engineering, Vol 17, pp. 734-749, 2005.
[2] B.S. Manjunath and P. Salembier and T. Sikora Introduction to MPEG-7: Multimedia Content Description Interface, Wiley, 2002.
[3] J. Posner and J. A. Russell and B. Peterson The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology, Development and Psychopathology, Vol. 17, pp 715–734, 2005.
[4] P. Ekman Basic Emotions, in Handbook of Cognition and Emotion, ed. T. Dalgleish and M. Power, Wiley, 1999.
[5] P. Viola, M. Jones Robust Real-time Object Detection, Second international workshop on statistical and computational theories of vision – modeling, learning, computing, and sampling, Vancuver, Canada, 2001.
[6] T. Chen, W. Yin, X. S. Zhou, D. Comaniciu, and T. Huang Illumination Normalization for Face Recognition and Uneven Background Correction Using Total Variation Based Image Models, Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition, 2005.
[7] Open computer vision library OpenCV web page http://sourceforge.net/projects/opencvlibrary/, accesed on 15.10.2008.
[8] G. Shakhnarovich, B. Moghaddam, Face Recognition in Subspaces, Handbook of Face Recognition, Eds. Stan Z. Li & Anil K. Jain, Springer-Verlag, 2004.
[9] I. Cohen, Q. Tian, X. Sean, Z. Thomas and S. Huang Feature selection using principal feature analysis, Proceedings of the 15th international conference on Multimedia, pp. 301 - 304, 2007.
[10] M. J.Escobar, J. R. del. Solar Biologically-based face recognition using Gabor filters and log-polar images, Int. Joint Conference on Neural Networks, pp-1143- 1147, 2002.
[11] S. Shan, P. Yang, X. Chen, W. Gao AdaBoost Gabor Fisher Classifier for Face Recognition, Conference AMFG, pp. 279-292, 2005.
[12] X. Wang, X. Tang, Bayesian face recognition using Gabor features, Proc. ACM SIGMM workshop on Biometrics methods and applications, pp. 70-73, 2003.
[13] X. He, and P. Niyogi Locality Preserving Projections, Advances in Neural Information Processing Systems 16 (NIPS), Vancouver, Canada, 2003.
[14] G. Guo, Y. Fu, C. R. Dyer Image-Based Human Age Estimation by Manifold Learning and Locally Adjusted Robust Regression, IEEE Transaction on Image Processing, Vol. 17, No. 7, July, 2008.
[15] J. Alon, V. Athitsos, Q. Yuan, S. Sclaroff A Unified Framework for Gesture Recognition and Spatiotemporal Gesture Segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 99, No. 1, 2008.
[16] L.-P. Morency and A. Quattoni and T. Darrell Latent-Dynamic Discriminative Models for Continuous Gesture Recognition, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Los Alamitos, CA, USA, 2007.
[17] T.-K. Kim and R. Cipolla Gesture Recognition Under Small Sample Size, Proc. Asian Conf. Computer Vision, pp. 335-344, 2007.
[18] R. Wimmer, P. Holleis, M. Kranz, A. Schmidt Thracker - Using Capacitive Sensing

for Gesture Recognition, In Proceedings of the 6th International Workshop on Smart Appliances and Wearable Computing (IWSAWC), Lisbon, Portugal, 2006.

[19] A. Poole, and L. J. Ball Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future Prospects, in Encyclopedia of Human Computer Interaction, IGI Global, 2005.

[20] M.-C. Su, K.-C. Wang, G.-D. Chen An eye tracking system and its application in aids for people with severe disabilities, Biomedical engineeringapplications, basis & communications, Vol. 18, No. 6, 2006.

[21] J. Tu. H. Tao, T. Huang Face as mouse through visual face tracking, Computer Vision and Image Understanding 108, pp. 35–40, 2007.

[22] G. Gonzalez, B. L´opez, J.L. de la Rosa The Emotional Factor: An Innovative Approach to User Modelling for Recommender Systems, Workshop on Recomendationa and Personalisation for e-Commerce, Adaptive Hypermedia, p´ag. 90-99, M´alaga, 2002.

[23] Gustavo Gonz´alez, Beatriz L´opez, Josep Ll. de la Rosa Managing Emotions in Smart User Models for Recommender Systems, Sixth Internacional Conference onEnterprise Information Systems (ICEIS 2004), vol. 3: 303-308, Porto (Portugal), April 2004.

[24] Hastie, T., Tibshirani, R., Friedman, J. The Elements of Statistical Learning, Springer, New York, USA, 2001.

[25] M. Haringer, S. Beckhaus Framework for the measurement of affect in interactive experiences and games, CHI’08, Florence, Italy, 2008.

Sadržaj: *Jedan od najznačajnjih nedostataka savremenih informaconih i komunikacionih tehnologija i uređaja predstavlja problem korisničkog interfejsa. U centru pažnje istraživača je da se pozabave koncepcijom personalizacije na bazi modeliranja korisničkih procedura. Glavna teškoća pri tome je zahtev za korisničkim modelom. Studije i praktično iskustvo pokazuju da u jednu ruku efikasnost korisničkog modeliranja zavisi od poznavanja podataka o korisniku, a sa druge strane za korisnika je često vrlo teško da obezbedi mogućnost pružanja podataka o svojim osećajima i ponašanju. Stoga postoji potreba za neinvazivnim prikupljanjem podataka. U tu svrhu postoji nekoliko različitih pristupa kao što su npr. distribuirani senzori i analiza videa u realnom vremenu, koja omogućuje identifikaciju specifične informacije o ponašanju korisnika.*

U ovom radu prikazan je problem utvrđivanja neinvazivne akvizicije podataka za korisničko modeliranje. Uz to je prikazana karakteristična arhitektura jednog takvog sistema. U tom kontekstu je uveden niz tehnika za obradu informacija. Predloženi pristup sadrži detalje akvizicije podataka na bazi video analize u realnom vremenu. Glavni deo rada je fokusiran na postprocesiranje podataka o korisniku, pri čemu se koristi geometrijska, topološka i logička analiza u cilju postizanja što boljeg korisničkog modela.

Ključne reči: *korisničko modeliranje, neinvazivna akvizicija podataka, fuzija podataka*

NEINVAZIVNA AKVIZICIJA PODATAKA
KOD KORISNIČKOG MODELIRANJA

Andrej Košir, Marko Tkalčič, Jurij Tasič