Automated Facial Pose Extraction From Video Sequences Based on Mutual Information

Georgios Goudelis, Anastasios Tefas, and Ioannis Pitas

***Abstract*—Estimation of the facial pose in video sequences is one of the major issues in many vision systems such as face-based biometrics, scene understanding for humans, and others. The proposed method uses a novel pose estimation algorithm based on mutual information to extract any required facial poses from video sequences. The method extracts the poses automatically and classifies them according to view angle. Experimental results on the XM2VTS video database and on a new database created for the needs of this research indicated a pose classification rate of 99.2% while it was shown that it outperforms a principal component analysis reconstruction method that was used as a benchmark.**

***Index Terms*—Biometrics, facial pose detection, mutual information.**

I. Introduction

FACIAL pose is one of the major issues concerning surveillance systems based on human behavior and intentions, as well as for face-based biometric applications. Facial pose estimation from video sequences is a task of great importance for vision systems performing scene understanding for human–computer interfaces or security surveillance [1]. A number of works can be found in the literature that attempt to estimate facial pose or to use this information for a number of different applications.

In [1], an analysis of face similarity distributions under varying head pose for different types of image transformation with the aim of understanding pose in similarity space is presented. In this work, the use of Gabor filters and principal component analysis (PCA) as transformation of prototypes images in order to emphasize pose differences is examined. In [2], a deformable graph is used to determine face position and pose from learned models. However, the method is highly time-consuming and is not appropriate for real-time applications. In [3], the work on eigenfaces is extended to modular eigenspaces in order to estimate the pose of a face, while in [4] the combination of support vector regression and modular support vector machines (SVMs) was used for pose estimation and face detection, respectively.

Other approaches use video in order to solve the pose estimation problem or to take advantage of pose extraction for applications such as face-based biometrics. More specifically, in [5], each registered person is represented by a low-dimensional appearance manifold in the ambient image space. This manifold is approximated by piecewise-linear subspaces, and the dynamics among them are embodied in a transition matrix learned from an image sequence. In [6], a method for real-time multiview face detection and facial pose estimation is described. The method employs a convolutional network to map face images to points on a manifold parameterized by pose and nonface images to points far from manifold. The network is trained by optimizing a loss function of three variables: image, pose, and face/nonface label. Finally, in [7], an independent component analysis (ICA)-based approach is presented for learning view-specific subspace representations of the object from multiview face examples. Two variants of ICA, namely independent subspace analysis (ISA) and topographic independent component analysis (TICA), take into account higher order statistics needed for object view characterization. ICA, TICA, and ISA are proven to learn view-specific basis components from the mixture data.

In this paper, we propose a novel method for automatic pose extraction in head-and-shoulder videos. The method is able to find any pose required. It is based on mutual information and evaluates the information content of each facial image (contained in a video frame) of facial poses in comparison to a given ground truth image. The experiments produced a pose classification rate of 99.2% on all examined pose cases.

The remainder of the paper is organized as follows. In Section II the algorithm for automated facial pose extraction is presented, followed by a short description of mutual information. In Section III, the description of the experimental procedure and the experimental results can be found. Section IV concludes the paper with a discussion.

II. Facial Pose Estimation

Mutual information (MI) has been previously used in computer vision, for example in image registration [8] or in audio-visual speech acquisition [9]. The mutual information of two random variables measures the mutual dependence of the two variables [10]–[12]. In our case, mutual information measures the dependency of the information contained in two video frames. The closer the mutual information between two frames is to zero, the less information one frame contains about the other and *vice versa*.

The underlying rationale of the MI for measuring the distance between facial poses in probabilistic terms is that it induces a certain structure in terms of "Information Geometry" [12]. Using the MI, the distance between facial poses is not measured in an Euclidean or a metric space but instead it is measured in a Riemmanian manifold [13]–[15] that is defined by the mutual dependence of facial poses. Under the assumption that the poses, and in general the distribution of faces, can be more safely modeled as a Riemmanian manifold (instead of an Euclidean

Manuscript received January 15, 2007; revised July 19, 2007. This work was supported in part by the Network of Excellence under project BioSecure IST-2002-507634 under Information Society Technologies (IST) priority of the 6th Framework Programme of the European Community. This paper was recommended by Associate Editor Y. Rui.

The authors are with the Aristotle University of Thessaloniki, Thessaloniki 55438, Greece (e-mail: goudelis@aiia.csd.auth.gr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2008.918457

space), the Kullback–Leibler (KL) divergence is the right metric for measuring distance. The use of such probabilistic measures (especially the KL divergence) has been very popular in nonnegative matrix factorization (NMF) methods [16], [17] for face reconstruction and face recognition. To the best of the authors' knowledge, this is the first work that uses the manifold that is defined by the MI distance in order to measure the similarity between facial poses. The method performed very well in our experimental procedure and appears to be robust against small changes in scale and illumination.

Let X be a discrete random variable with a set of possible outcomes $\mathcal{A}_X = \{a_1, a_2, \ldots, a_N\}$ having probabilities $\{p_1, p_2, \ldots, p_N\}$, with $p_X(x = a_i) = p_i$, $p_i \geq 0$, and $\sum_{x \in \mathcal{A}_X} p_X(x) = 1$. Entropy measures the information content or "uncertainty" of X and it is given by [18]

$$H(X) = -\sum_{x \epsilon \mathcal{A}_X} p_X(x) \log p_X(x). \tag{1}$$

The *joint entropy* is a statistic measure that summarizes the degree of dependency of random variable X on random variable Y. The *joint entropy* of X, Y is expressed as

$$H(X,Y) = -\sum_{x,y \in \mathcal{A}_X, \mathcal{A}_Y} p_{XY}(x,y) \log p_{XY}(x,y) \tag{2}$$

where $p_{XY}(x,y)$ is the joint probability density function. For two random variables X and Y, the *conditional entropy* of Y given X is denoted as $H(Y|X)$ and is defined as

$$\begin{aligned} H(Y|X) &= \sum_{x \in \mathcal{A}_X} p_X(x) H(Y|X = x) \\ &= -\sum_{x,y \in \mathcal{A}_X, \mathcal{A}_Y} p_{XY}(x,y) \log p_{XY}(x|y) \end{aligned} \tag{3}$$

where $p_{XY}(x|y)$ denotes the conditional probability. The conditional entropy $H(Y|X)$ is the uncertainty in Y, given knowledge of X. It specifies the amount of information that is gained by measuring a variable when already knowing another one. It is very useful if we want to know whether there is a functional relationship between two data sets (e.g., two facial image regions). The MI between the random variables X and Y is given by

$$I(X,Y) = -\sum_{x,y \in \mathcal{A}_X, \mathcal{A}_Y} p_{XY}(x,y) \log \frac{p_{XY}(x,y)}{p_X(x) p_Y(y)} \tag{4}$$

and measures the amount of information conveyed by X about Y. The relation between the mutual information and the joint entropy of random variables X and Y is given by

$$I(X,Y) = H(X) + H(Y) - H(X,Y) \tag{5}$$

where $H(X)$ and $H(Y)$ are the marginal entropies of X and Y. The MI is a measure of the additional information known about X when Y is given as

$$I(X,Y) = H(X) - H(X|Y) \tag{6}$$

where $H(X)$ is the marginal entropy, $H(X|Y)$ is conditional entropy, and $H(X,Y)$ is the joint entropy of X and Y. According to (5), the MI provides us with a measure of correspondence between X and Y. We can also see from (6) that the MI is reduced, if X carries no information about Y.

A. Calculation of MI Between Video Frames

In order to calculate the MI between two video frames, we have to use a reference frame and a test frame denoted by X and Y accordingly, having L pixels each. We can consider pixel values as outcomes $\mathcal{A}_X = \{a_1, a_2, \ldots, a_L\}$ of a random variable. The probabilities p_X and p_Y in (4) are estimated by the histograms of the images U and V, while the joint probability p_{XY} is estimated by the joint histogram of the images [19]. The density probabilities of both images are estimated using the Parzen Window technique [20], [21]. This is a classical technique used in neural networks for estimating a probability density function (pdf) from a sample.

While entropy for an image remains fixed, joint entropy and MI of two images vary as the 1-1 correspondence between the pixels from each image changes with every geometrical alignment. When MI is maximized, the geometric relationship, under which one image explains the other most effectively, is achieved. In other words, the maximization of MI provides image registration. Mutual information has the following properties [22].

- It is symmetric: $I(X,Y) = I(X,Y)$. However, although it is a logical property in theory, MI is not symmetric in practice. Many aspects of a registration method, such as interpolation and number of samples, can result in differences in outcome when registering X to Y or Y to X.
- $I(X,X) = H(X)$. The information image X contains about itself is equal to the information (entropy) of image X.
- $I(X,Y) \leq H(X)$, $I(X,Y) \leq H(Y)$. The information the images contain about each other can never be greater than the information contained in the images themselves.
- $I(X,Y) \geq 0$. The uncertainty about X cannot be increased when knowing about Y.
- $I(X,Y) = 0$, if and only if X and Y are independent. When X and Y are not in any way related, no knowledge is gained about one image when the other is given.

1) Computational Complexity: The computational complexity of MI consists of calculating three histograms for each of the three color components R, G, and B of two frames. If the frame size is n pixels then for one histogram n additions are required. At first, three histograms are calculated which consist of $3n$ additions. We also need another $3N^2$ multiplications, $(N-1)^2$ additions, and N^2 logarithmic calculations for computing (4), where N is the number of the video sequence gray levels, separately for each of the three color channels. Thus, for calculating the MI between two frames, $O(n + N^2)$ additions, $O(N^2)$ multiplications, and $O(N^2)$ logarithmic calculations are required.

B. Pose Classification

Most videos are obtained using an uncalibrated camera. This means that we do not know the three angles that define a person's pose with respect to the camera reference system. Therefore, we consider pose estimation as a classification problem in which we assign a facial image to a particular pose class (e.g., frontal, left/right profile) by examining its similarity

Fig. 1. Sample of XM2VTS video sequence. The person starts from the frontal pose, turn his or head head from right to left, and returns to the frontal pose.

with other images of the video. In order to describe how the pose is assigned to a class, let us describe the enrolment procedure for a candidate reference person r. The system contains an image database $\mathcal{C} = \bigcup_r C_r$, where $\mathcal{C}_r$ is the set of images assigned to the reference person r. Suppose that $\mathcal{C}_{jr}$ is subset of the $\mathcal{C}_r$ that contains the images of jth pose of person r so that $\mathcal{C}_r = \bigcup_j \mathcal{C}_{jr}$. Let the video V_r be partitioned to the set of N frames $F_{1r}, \ldots, F_{Nr}$. Each video frame F_{kr} (where $k = 1, \ldots, N$) is examined and compared with the images of the set $\mathcal{C}_r$ in order to assign it to a pose set $\mathcal{C}_{jr}$. The mean MI

$$I_m(F_{kr}, \mathcal{C}_{jr}) = \frac{1}{\mathcal{N}(\mathcal{C}_{jr})} \sum_{\mathbf{c}_i \in \mathcal{C}_{jr}} I(F_{kr}, \mathbf{c}_i) \tag{7}$$

is calculated using (6) for every $F_{kr} \in V_r$. $\mathbf{c}_i$ is an image of $\mathcal{C}_{jr}$ and $\mathcal{N}(C_{jr})$ denotes the cardinality of the set C_{jr}. The frame F_{kr} is assigned to the pose class corresponding to the maximum $I_m(F_{kr}, \mathcal{C}_{jr})$ as

$$pose = \arg\max_j I_m(F_{kr}, \mathcal{C}_{jr}). \tag{8}$$

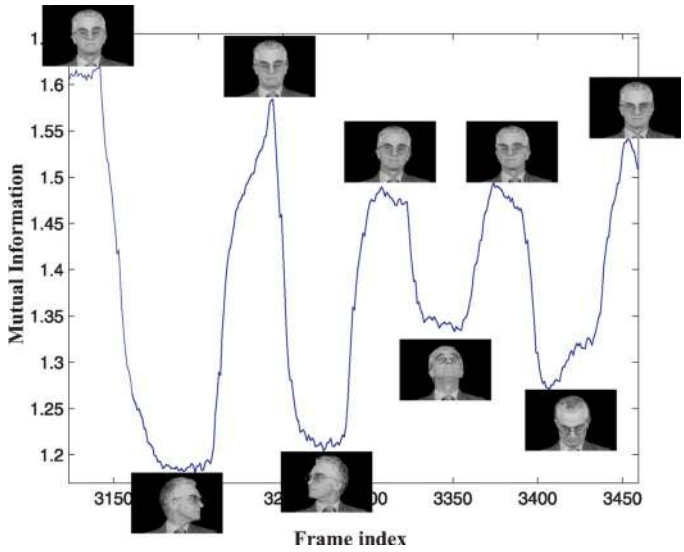

Fig. 2. Mutual information plot for a smooth head movement vs video frame time index. The characteristic facial image poses are superimposed.

III. Description of Experiments and Results

In order to make the procedure more clear to the reader and present a possible use of the proposed method, we set forth a hypothetical scenario that describes a verification system based on the proposed face detection algorithm, as a real-world application.

A surveillance camera is installed in a building where special security issues are required and comprise part of the security system. In the specific system, a number of persons have been enrolled. During the enrollment procedure, the person is asked to turn his or her head in all possible directions in front of a recording camera in order to store in a database every facial pose of the person. For some poses that seem to be rich in information content like frontal, right, and left profiles, a pose-specific training procedure is followed. This way, all of the crucial facial poses of a client are learned by the system. Let us suppose that a person is requiring access to the building (i.e., an identity claim occurs). Unlike the enrollment procedure which is supervised, the testing procedure is fully automatic. The surveillance camera is recording the scene while a face detector locates the facial area in every frame. The pose classes are used according to the recognition/verification algorithm that is used, and a decision is made for the acceptance or the rejection of the claim.

A. Evaluation of the Pose Detection Algorithm

For testing the capability of the algorithm having many persons and different video sessions of them, we used the XM2VTS video database. This database contains four recordings of 295 subjects taken over a period of four months. Each recording is comprised of a speaking head shot and a rotating head shot. Sets of data taken from this database are available including high-quality color images, 32-kHz 16-bit sound files, video sequences, and a 3-D model. In the first shot of each session, the person reads a given text, while in the second shot each person moves its head in all possible directions allowing multiple views (poses) of the face.

In our framework, video is processed frame by frame. Only the subsampled luminance is used in order to reduce processing

Fig. 3. Successful pose estimations. (a) Ground truth images. (b) Results produced.

Fig. 4. Example of the type of poses used for the evaluation of the pose detection algorithm.

time. Afterwards, the uniform background is removed using a grassfire algorithm [23]. To achieve a better and more accurate verification rate, the algorithm resizes each video, according to a factor produced by a given standard distance between the right and left eyes, when the subject is in frontal position, while it keeps the frame size stable. This way, the scaling problem occurring in different sessions of the same person is resolved (at least partially), while the head is aligned for the frame representing the frontal face and, consequently, for the most part of the remaining of the movement. It should be noted that the eyes are automatically detected using the method described in [24]. A sample sequence of already processed frames is illustrated in Fig. 1.

For each of the persons examined, a ground truth image representing the required pose is used. This image is always taken from a different session from the one examined. The ground truth constitutes F_{kr} in (7), while $\mathbf{c}_i$ is every following frame of the examined video input. This way, each frame is compared with the ground truth and their mutual information is stored in a vector $\mathbf{q}$.

The plot of this vector entries $\mathbf{q}_i$, (for $i = 1, \ldots, N$) for a smooth movement is given in Fig. 2. The mutual information in this case has been calculated using the frontal face of the person taken from a different session than the one under examination.

The plot clearly shows how mutual information changes, as the head passes from pose to pose. The mutual information is maximized when the head pose is close to the frontal position. This clearly shows how the MI "detects" the similarity between the first frame which represents the frontal face and every near to frontal pose that appears within the video sequence.

1) Experiments on Aligned Scaled Frames: Likewise, we repeat the same procedure using specific poses extracted from a different video session than the test one. The method was tested on 120 different persons, each time using (as ground truth) images taken from each of the four different sessions. Thus, 4320 (4 sessions $\times$ 3 poses $\times$ 3 tests to every session$\times 120$ persons) testings where carried out. The result obtained was of high accuracy showing that the algorithm was able to find correctly 99.2% of the required poses. Some successful matchings are presented in Fig. 3.

2) Experiments on Nonaligned Artificially Partially Occluded Frames: To see how MI reacts when the subject is partially occluded and how a face tracker affects the performance, we performed the above experiment on the same database but having the videos artificially occluded. At this point, it should be noted that, during this experiment, no eye detector information was used and, consequently, no alignment of the faces or rescaling has taken place. The initial frame size was of 216×144 pixels and, for every following experimental cycle, a fixed-size bounding box has used which gradually was reduced to the face area. An example of the area participating in every experimental cycle is illustrated in Fig. 5. The experiment has run for up, down, left, right, medium left, and medium right poses. As reference poses, only nonoccluded face images from a different session have been used. An example of the type of poses used for the experiment is presented in Fig. 4. The bounding box size was fixed at 216×144, 159×109, 123×91, 87×77, 77×69, and 30×20 pixels, for every experimental cycle, respectively. The results of this experiment are shown in Fig. 6. At this point, it should be noted that the best results where obtained when the tracker bounding box was fixed on 87×77 and 77×69 pixels where the area examined contains only facial information. The grate decrease obtained for the following dimensions is due to the small area examined which in every case it could contain different parts of the face (e.g., an eye, a part of the nose, or the mouth).

3) Experiments on Nonaligned Frames Containing a Complicated Background: In order to reinforce our experimental re-

Fig. 5. Example of XM2VTS video database subject under artificial occlusion. The gradual decrease of the frame area examined represents the different sizes of tracking bounding boxes applied.

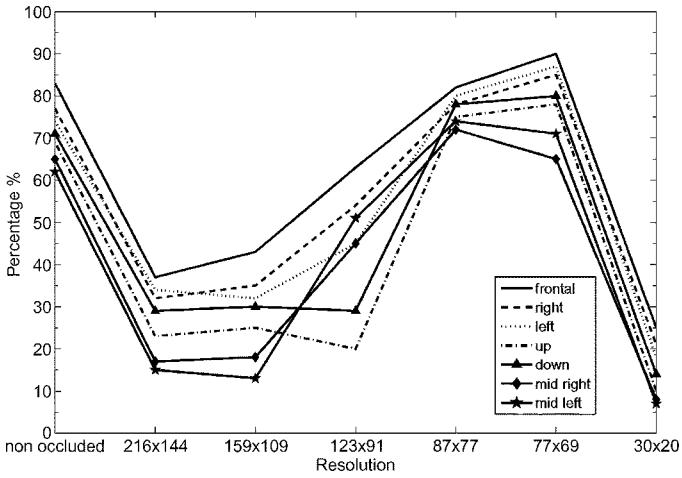

Fig. 6. Results on XM2VTS video database artificially partially occluded.

Fig. 7. Data samples used for the experimental procedure. (a) Outdoor session. (b) Indoor session.

sults and show in more detail how the MI algorithm performs, we have also performed experiments in a database especially produced for the needs of the specific research. The database consists of 45 individuals and two sessions each. Every subject has been recorded indoors and outdoors while the background in both cases is complicated. In both sessions, the subjects have been asked to rotate their heads in an XM2VTS video database manner. The initial size of the videos was of 720×576 pixels. A few subjects participating in the database are shown in Fig. 7. It is worth noting here that the difference of the illumination between the two sessions is distinguished.

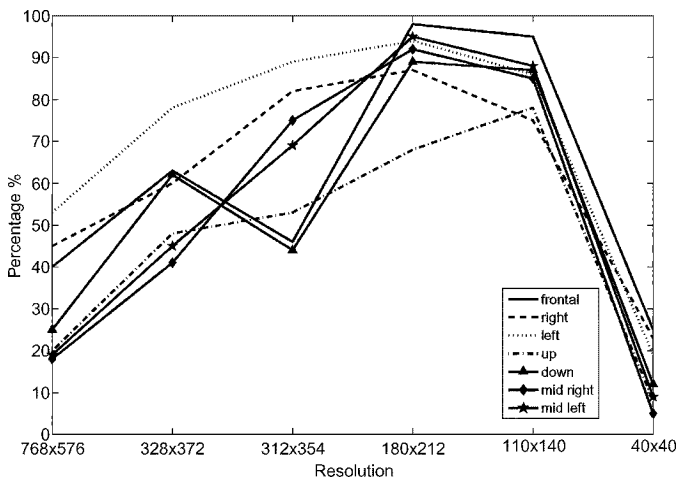

Fig. 8. Percentages of successful facial pose detections for different sizes of the tracker bounding box.

More specifically, we wanted to evaluate the ability of mutual information to detect a pose within a video sequence when the background is complicated and to see how the performance changes with the use of a face tracker. For this reason, we performed the following experiments; working on the database described above and by using the whole frame, we applied the pose detection algorithm using reference frames as ground truths extracted from the indoor shots trying to estimate the pose in the outdoor shots and vice versa. In following, by using face tracker information with a fixed size bounding box, the same procedure was followed while for every experimental cycle the bounding box was gradually shrinking to the face area. Starting by using the initial size of the image (768×576) the bounding box was fixed to 328×372, 312×354, 180×212, 110×140 and 40×40 pixels for every experimental cycle respectively. The results are illustrated in Fig. 8. A few example plots of the vector $\mathbf{q}$ that stores the mutual information value are presented in Fig. 9. All of the examples represent cases where face tracker of various sizes bounding boxes have been used. It can be seen that when tracker information is used, mutual information acquires its high

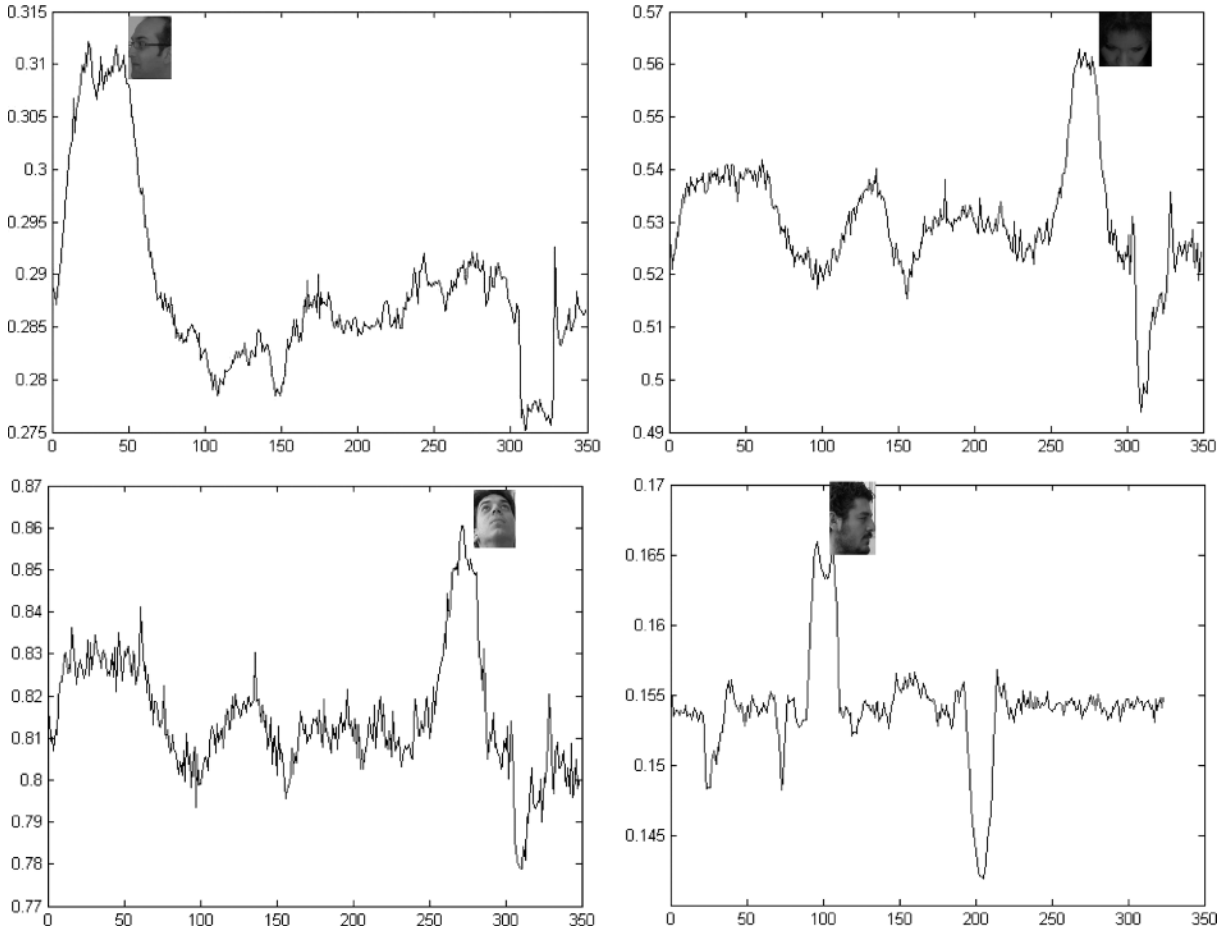

Fig. 9. Mutual information plots for random examples where face tracker has been used. The MI value difference obtained in every case is due to the use of different sizes of bounding boxes, respectively.

TABLE I
PERCENTAGES (%) OF CORRECT FACIAL POSE CLASSIFICATION

Method	Right	Mid-Right	Frontal	Mid-Left	Left	Up	Down
PCA on XM2VTS	75.2	63.3	85.5	65.7	76.0	79.7	81.6
Proposed on XM2VTS	85.1	65.3	90.2	71.4	87.1	78.3	80.1
PCA on our database	73.4	69.5	83.9	68.2	74.9	78.4	87.7
Proposed on our database	87.2	91.9	98.1	94.8	94.1	68.0	89.1

values exclusively when the pose is presented in the video sequence. As obtained in the previous experiment, the best results were given for a specific dimension of the tracker bounding box where the face area was well described.

4) Comparison of Pose Detection Algorithm With PCA Reconstruction Method: To make a comparison of the pose detection algorithm with a method that could be used as a benchmark, we used a PCA reconstruction method for pose classification. For each pose class, a PCA model was constructed. The model with the smaller reconstruction error was the one finally classified (i.e., smallest L_2 norm distance). This scenario is like using eigenfaces method for every pose and when an unknown pose arrives, the test image is projected to all different pose-subspaces. The one obtaining the minimum L_2 distance is the winner. The experiment was performed for both the above described databases. PCA was trained with ground truth facial pose images extracted by the video sequences. The experiment has run on different sessions for different bounding box dimensions and no alignment or re-scaling has taken place. Due to space limitations we report only the best results for both methods produced by the use of bounding boxes with sizes 77 $\times$ 69 and 180 $\times$ 212 pixels, for the XM2VTS and our database respectively. The results of the comparison of the two methods, are given in Table I.

It can be easily seen that the performance of the proposed method is significantly better for almost every pose examined. On average for both databases, the proposed method outperforms PCA reconstruction method by 7.89%.

IV. CONCLUSION

In this paper, a novel way for automatic facial pose extraction on MI is proposed. The information in video frames is compared with the information contained in a ground truth image representing the required pose. Experimental results on the XM2VTS video database, as well as on a new video database created

for the needs of this research that involves complicated backgrounds, show that the algorithm is able to perform very well for a number of different requested poses when tracker information is used. The method proved to outperform a PCA reconstruction method which was used as a benchmark. It is worth noting that the pose detection algorithm proved to be robust in small variations of scale and illumination.

REFERENCES

[1] J. Sherrah, S. Gong, and E. Ong, "Face distribution in similarity space under varying head pose," *Image Vis. Comput.*, vol. 19, no. 11, 2001.

[2] N. Kruger, M. Potzsch, and C. von der Malsburg, "Determination of face position and pose with a learned representation based on labeled graphs," *Image Vis. Comput.*, pp. 665–673, 1997.

[3] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Seattle, WA, Jun. 1994, pp. 1197–1215.

[4] Y. Li, S. Gong, and H. Liddell, "Support rector regression and classification based multi-view face detection and recognition," in *Proc. 4th IEEE Int. Conf. Automatic Face Gesture Recognition*, 2000, p. 300.

[5] K.-C. Lee, J. Ho, M.-H. Yang, and D. J. Kriegman, "Video-based face recognition using probabilistic appearance manifolds," in *CVPR*, 2003, pp. 313–320.

[6] R. Osadchy, M. Miller, and Y. LeCun, "Synergistic face detection and pose estimation with energy-based model," in *Proc. Adv. Neural Inf. Process. Syst.*, 2005, pp. 84–91.

[7] S. Z. Li, X. Lu, X. Hou, X. Peng, and Q. Cheng, "Learning multiview face subspaces and facial pose estimation using independent component analysis," *IEEE Trans. Image Process.*, vol. 14, no. 6, pp. 705–712, Jun. 2005.

[8] P. A. Viola, Alignment by Maximization of Mutual Information Tech. Rep. AITR-1548, 1995, p. 156.

[9] D. Roy and A. Pentland, "Learning words from natural audio-visual input," in *Proc. Int. Conf. Spoken Language Process.*, Sydney, Australia, 1999, vol. 4, pp. 1279–1283.

[10] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, Jul. 1948.

[11] C. Shannon and W. Weaver, *Mathematical Theory of Communication*. Chicago, IL: Univ. of Illinois Press, Jun. 2002.

[12] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed. New York: McGraw-Hill, 1991.

[13] O. P. Ferreira and B. F. Svaiter, "Kantorovich's theorem on Newton's method in Riemannian Manifolds," *J. Complex.*, vol. 18, no. 1, pp. 304–329, 2002.

[14] E. S. Polovinkin, "Riemannian integral of set-valued function," in *Optimization Techniques*, 1974, pp. 405–410.

[15] A. M. Bloch and P. E. Crouch, *Nonholonom. Control Syst. Riemannian Manifolds*, vol. 33, no. 1, pp. 126–148, Jan. 1995.

[16] S. Zafeiriou, A. Tefas, I. Buciu, and I. Pitas, "Exploiting discriminant information in nonnegative matrix factorization with application to frontal face verification," *IEEE Trans. Neural Netw.*, vol. 17, no. 3, pp. 683–695, Jun. 2006.

[17] D. Lee and H. Seung, "Algorithms for non-negative matrix factorization," in *Proc. NIPS*, 2000, pp. 556–562 [Online]. Available: citeseer.ist.psu.edu/lee01algorithms.html

[18] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley-Interscience, 1991.

[19] P. Thevenaz and M. Unser, "An efficient mutual information optimizer for multiresolution image registration," in *Proc. IEEE Int. Conf. Image Process.*, Chicago, IL, Oct. 4–7, 1998, vol. I, pp. 833–837.

[20] E. Parzen, "On the estimation of a probability density function and mode," *Ann. Math. Stat.*, vol. 33, pp. 1065–1076, 1962.

[21] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.

[22] J. Aczel and Z. Daroczy, *On Measures of Information and Their Characterizations*. New York: Academic, 1975.

[23] C. Kotropoulos, A. Tefas, and I. Pitas, "Morphological elastic graph matching applied to frontal face authentication under well-controlled and real conditions," *Pattern Recognit.*, vol. 23, no. 12, pp. 31–43, Oct. 2000.

[24] S. Asteriadis, N. Nikolaidis, I. Pitas, and M. Pardas, "ADetection of facial characteristics based on edge information," in *Proc. 2nd Int. Conf. Comput. Vis. Theory Appl.*, Barcelona, Spain, Mar. 2007, pp. 247–252.