Proc. of Int. Conf. on Advances in Information Technology and Mobile Communication 2013

Face Pose Classification Method using Image Structural Similarity Index

Hari C.V. and Praveen Sankaran
Department of Electronics & Communication Engineering
National Institute of Technology Calicut, Calicut, Kerala, India - 673 601
Email:haricv@gmail.com, psankaran@nitc.ac.in

Abstract—**Face pose estimation methods try to identify/ classifythe position and orientation of human faces present in an image.This paper proposes a new method of face pose classificationbased on the structural similarity index. It is based on themeasure of the similarity between the facial image with a facialpose with a set of images in the database with different poses.**

Index Terms—**Face Pose Classification, Structural Similarity Index(SSIM)**

I. Introduction

Face pose classification is one of the major research areasin the field of computer vision and robotics. Human computerinterface, driver assistance. . . etc are some of the major applicationsof pose estimation. Pose estimation is the process ofidentifying the position and orientation of a human face withrespect to a reference coordinate system.

Chutorian and Trivedi [1] gives a survey on different headpose identification methods. According to them, head poseestimation is the process of inferring the orientation of ahuman head from digital imagery as in a computer visioncontext. Almost all of the head pose estimation methodsare based on some rigid model with inherent limitations.Head pose estimation problem is complicated with varyingconditions like lighting, background and camera geometry.

Niyogi and Freeman [2] proposed a head pose identificationmethod based on a non-linear mapping from the input imageto an output parametric description. This mapping throughthe examples from a training set, which gives the pose asthe nearest neighbour of the input. It is a modified vectorquantization which stores an output parameter code with eachquantized input code with a tree structured vector quantizerfor efficient indexing.

Beymer [3] proposed pose estimation by finding the eyesand nose lobe features. This method is also template-based,with tens of facial feature templates covering different posesand different people. The recognizer also applies an affinetransform to the input to bring the three feature points intocorrespondence with the same points on the model.

Sherrah et.al. [4] proposed pose estimation with a similarity-to-prototypes philosophy. The similarities are calculated robustlywith the help of principal component analysis (PCA)which provide an identity invariant representation. Here, thedifferences in poses are enhanced by the orientation selectiveGabor filters. Different filter orientations are claimed to beoptimal at different poses.

More recently, Goudeliset. al. [5] proposed a methodfor automatic pose extraction in head-and-shoulder videos.Mutual information between the frames is the key idea behindthis technique. Mutual information evaluates the informationcontent of each facial image (contained in a video frame) offacial poses in comparison to a given ground truth image. Thismethod is able to find any pose required with a good accuracy.

In this paper, we extend the idea of image quality comparisonwith a ground truth image and propose a new facepose classification technique from a video sequence containinghead-and shoulder videos. This method is based on the structuralsimilarity of facial image from the input video (whichforms the test input image) to the facial images in the database(which are the ground truth images for the different classes).

II. Image Similarity Measurement

Common image similarity measuring techniques include Mean Square Error (MSE), Peak Signal to Noise Ratio(PSNR), Root Mean Square Error (RMSE), Mean AbsoluteError (MAE) and Signal to Noise Ratio (SNR). Another typeof image quality measurement is based on the models ofHuman Visual System (HVS) [6]. Structural similarity index (SSIM) [7] is a full reference metric, in other words, the measuring of image quality based on an initial uncompressed or distortion-free image as reference. SSIM was designed toimprove on traditional methods like peak signal-to-noise ratio(PSNR) and mean squared error (MSE), which have provedto be inconsistent with human eye perception. Since SSIM provides a comparative similarity between two images as shown in Fig.(1), its use can be expanded beyond the intended image quality analysis. We propose a novel SSIM based face pose classificationmethod that takes into consideration the fact that closer posesshould yield more similar images and hence should providehigher SSIM values.

A. Universal Image Quality Index

Universal Image Quality Index [8] measures the similarityas a factor of distortion using a combination of three differentfactors: loss of correlation, luminance distortion and contrastdistortion. If $X = \{X_i| = 1, 2, \ldots, N\}$ and $Y = Y_i| = 1, 2, \ldots, N\}$ be the original and the test signals respectively, the Universal Image Quality Index Q between X and Y isdefined as

DOI: 03.LSCS.2013.2.81

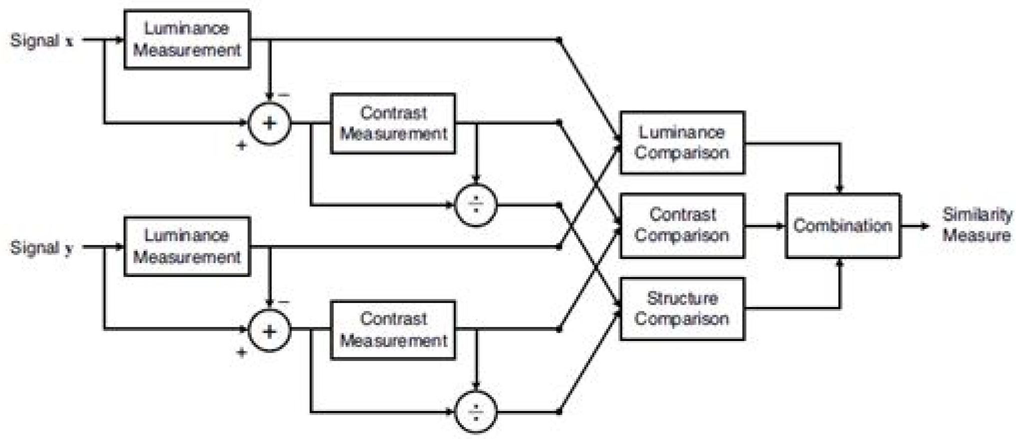


Fig.1. SSIM Measurement System [7]

$$Q = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \frac{2\mu_x \mu_y}{(\mu_x^2+\mu_y^2)} \frac{2\sigma_x \sigma_y}{(\sigma_x^2+\sigma_y^2)} \quad (1)$$

where,

$$\mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i, \ \mu_y = \frac{1}{N}\sum_{i=1}^{N} y_i$$

$$\sigma_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu_x)^2, \sigma_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \mu_y)^2$$

$$\sigma_x^2 = \frac{1}{N-1}\sum_{k=1}^{N}(x_i - \mu_x)(y_i - \mu_y)$$

In the Eq.1, the first, second and third components measuresthe linear correlation (correlation coefficient), luminancecomparison and contrast comparisons between X and Yrespectively. The dynamic range of correlation coefficient is[-1, 1] and for luminance and contrast comparisons it is [0, 1].So the dynamic range of Q will be [-1, 1].When applying this technique to images, the image qualityindex can be measured from the local regions of the imageusing a sliding window approach. The sliding window of a particular size is moved horizontally and vertically through the image; starting from the left corner of the image. Theimage quality index Q will be,

$$Q = \frac{1}{M}\sum_{j=1}^{M} Q_j \quad (2)$$

where M is the total number of steps and Q_j is the localquality index at j^{th} step.

B. Structural Similarity Index (SSIM)

Structural Similarity Index (SSIM) is a modified version ofUniversal Quality Index. Like Universal Quality Index, theSSIM value will be high for similar frames/images. SSIMindex collectively measure the changes in the luminance,contrast and structure between two signals or images.

$$S(x, y) = f(l(x, y), c(x, y), s(x, y)) \quad (3)$$

where$l(x, y)$ is the luminance comparison, $c(x, y)$ is thecontrast comparison and $s(x, y)$ is the structural comparison.From the Eq.3, it is clear that the SSIM index makes threecomparisons (luminance, contrast and structural) for similaritymeasurements.They are,

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2+\mu_y^2+C_1} \quad (4)$$

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2+\sigma_y^2+C_2} \quad (5)$$

$$s(x,y) = \frac{2\sigma_{xy} + C_3}{\sigma_x\sigma_y+C_3} \quad (6)$$

where $C_1 = (k_1L)^2$, $C_2 = (k_2L)^2$ and $C_3 = \frac{C2}{2}$ are small constants; L is the dynamic range of the pixel values, and K1 << 1 and K2 << 1 are scalar constants.

Finally by combining all the above mentioned comparisons, the Structural Similarity Index is given by,

$$S(x,y) = [\, l(x,y)]^{\alpha}\, [c(x,y)]^{\beta}\, [s(x,y)]^{\gamma} \quad (7)$$

where $\alpha > 0$, $\beta > 0$ and $\gamma > 0$ are parameters used toadjust the relative importance of the three components. Thesimplified SSIM measure is obtained by substituting $\alpha = \beta = \gamma = 0$ and $C3 = \frac{C_2}{2}$, then

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy}+C_2)}{(\mu_x^2+\mu_y^2+C_1)(\sigma_x^2+\sigma_y^2+C_2)} \quad (8)$$

Eq.8 is similar to the Eq.1 when C1 = C2 = 0. i.ethe Universal Quality Index is a special case of SSIM whenC1 = C2 = 0. But, the Universal Quality Index will produceunstable results when either $(\mu_x^2+\mu_y^2)$ or $(\sigma_x^2+\sigma_y^2)$ is zero.

DOI: 03.LSCS.2013.2.81

III. Face Pose Classification

Face pose identification is considered as a classificationproblem based on the SSIM value between a test image andreference ground truth images from multiple face-pose classes.Face pose is the relative orientation or angle of the face withthe camera. The three degrees of freedom of a human face canbe described as pitch, roll and yaw as shown in Fig.2.

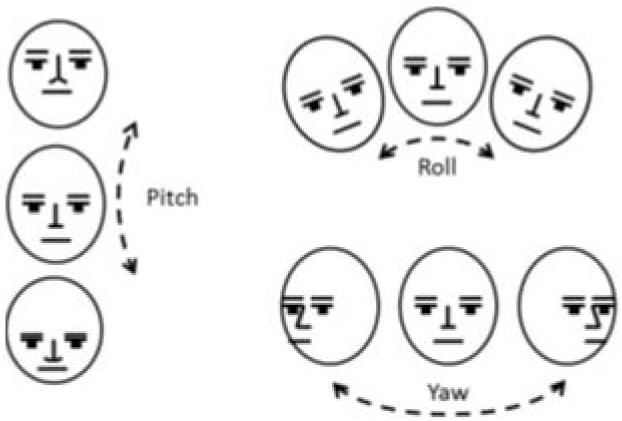


Fig.2. Face Pose Angles - pitch, roll and yaw [9]

SSIM value of the current frame/image is compared withall other face images in the database and is assigned to theclass of the image which provides the maximum SSIM value.For calculating the SSIM value, the system contains an imagedatabase, which has reference face poses of this particular person. The input video will be subdivided into differentframes and each frame's SSIM value will be calculated with allthe images of the database. The current frame of the referenceperson is assigned to the pose class corresponding to themaximum SSIM value of the image from the database.

IV. Experimental Results

SSIM based facial pose classification algorithm is implementedin MATLAB and tested in the Pointing'04 [10]database. This database consists of 15 sets of images. Eachset contains of 2 series of 93 images of the same person atdifferent poses. There are 15 people in the database, wearingglasses or not and having various skin color. The pose orhead orientation is determined by the angles varying from -90degrees to +90 degrees. Only luminance component is usedhere in order to reduce processing time.

The input video is sub-sampled into its frames. Each frameis compared with all the images of the database to calculate itsSSIM value for all frames. High SSIM value will be obtainedwhen the current frame and an image in the database will besimilar to each other. The corresponding image in the databasewill give the current pose of the input frame.

In order to describe how the pose is assigned to a class, letus describe the face pose identification method for a personin the database. Let us consider frame number 50 of the inputvideo which is shown in Fig.3 (a).

SSIM values which lie between frame 50 and all the images

a) Frame number 50 (b) Detected Pose from the Database

Fig.3. SSIM - Output for frame number 50

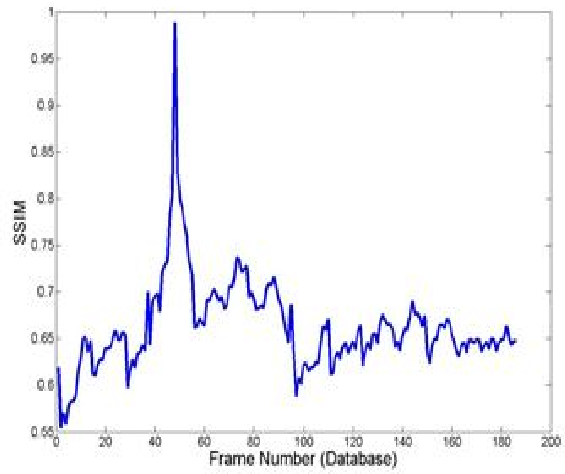


Fig. 4. SSIM Value of Frame number 50

in thedatabase is calculated. Image representing a specified pose in the database with the highest SSIM value will be the pose of frame 50 as shown in Fig. 3(b).

As shown in Fig.4, SSIM value of frame number 50 withall other reference images in the database is varying. SSIMvalue is maximum for the 50th image in the database whichis the same as the current frame under test. That image in thedatabase represents the pose of the current frame 50 of ourinput video.

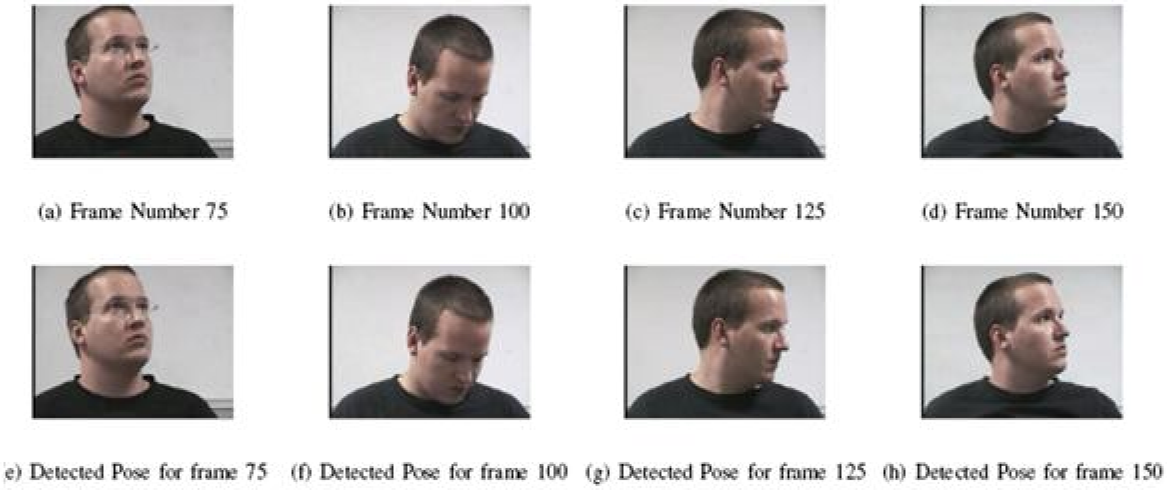
In this manner, we can identify any of the pose from theinput video. The inputs and outputs for frame numbers 75,100, 125 and 150 are as shown in the Fig. 5. SSIM values between the test frame and the images in the database are asshown in the Fig.7. From this graph it is clear that SSIM value is maximum for the image containing in the database whichhas same pose as in the input video and also that the SSIMvalue varies between 0 and 1.

A. Performance Analysis

If the number of sample images in the database is reduced,i.e. if all the poses are not available to us as ground truths, thenthe detected pose will be a pose nearest to the correspondingpose in the test input frame. For analyzing the performanceexperimentally, generate 'X' random numbers with uniform distribution. Corresponding numbered frames are left out from the training set and are used as frames for testing. This is repeated using different 'X' values. The 'X' values used hereare 5, 10, 25, 50 and 100. SSIM values for this experiment are as show in Fig.8, 9, 10, 11 and 12. From this figures,it is clear that, the confidence for the detection is reducing,because by reducing the number of frames from the database, more frame's SSIM value will be close together.

The confidence level for the pose classification reduces byleaving out frames from the database. Confidence level is

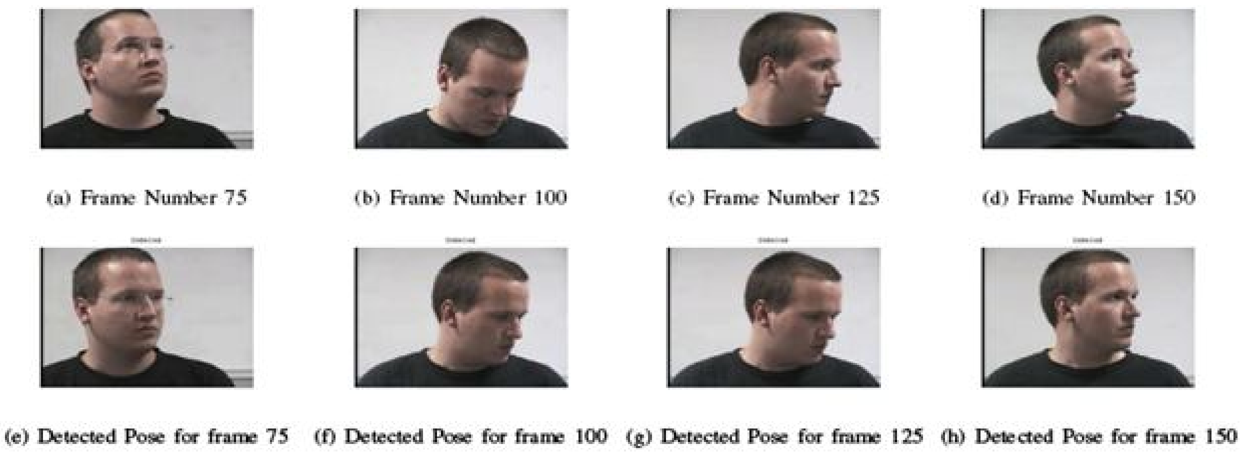
DOI: 03.LSCS.2013.2.81



(a) Frame Number 75 (b) Frame Number 100 (c) Frame Number 125 (d) Frame Number 150

(e) Detected Pose for frame 75 (f) Detected Pose for frame 100 (g) Detected Pose for frame 125 (h) Detected Pose for frame 150

Fig.5.Different frames and their detected poses



(a) Frame Number 75 (b) Frame Number 100 (c) Frame Number 125 (d) Frame Number 150

(e) Detected Pose for frame 75 (f) Detected Pose for frame 100 (g) Detected Pose for frame 125 (h) Detected Pose for frame 150

Fig. 6. Different frames and their detected poses in a reduced database (Leaving 10 frames from the database)

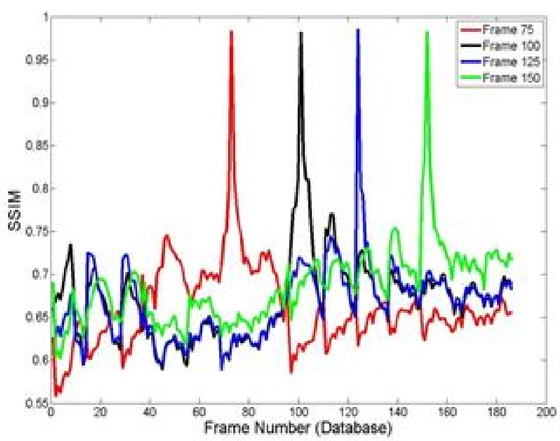


Fig.7. SSIM Value of Frame number 75,100,125 and 150

measured by taking the average of the ratio between peak valuesto the non-peak values. The confidence level measurement fordifferent databases is as shown in the Table. I. From this, it is clear that, the confidence level for detection is decreasing byreducing the number of images in the database.

The outputs for different frames for this reduced database are as shown in Fig.6 and Fig.13. There are 10 frames removed from the database in both cases and SSIM value will bemaximum for a pose that is closest to the input test posesince the original pose is not available in the database. Thedetected pose for frame number 50 is shown in Fig. 13. Here,the detected pose is not exactly similar with the input frame. The SSIM value is maximum for some other image and isshown as the detected pose of this frame. SSIM values offrame 50 for reduced database are as shown in the Fig.14. The inputs and outputs for frame numbers 75, 100, 125 and 150are as shown in the Fig. 6 with a reduced number of imagesin the database.

Table I: Confidence Level Measurement

Number of Training images reduced from database	Confidence measure(Peak/ Non-peak Average)	Time (Sec)
5	1.4588	11.68
25	1.3336	10.30
50	1.3116	8.70
75	1.3005	7.04
100	1.2508	5.54

V. Conclusion

This paper proposed a novel approach for face pose

DOI: 03.LSCS.2013.2.81

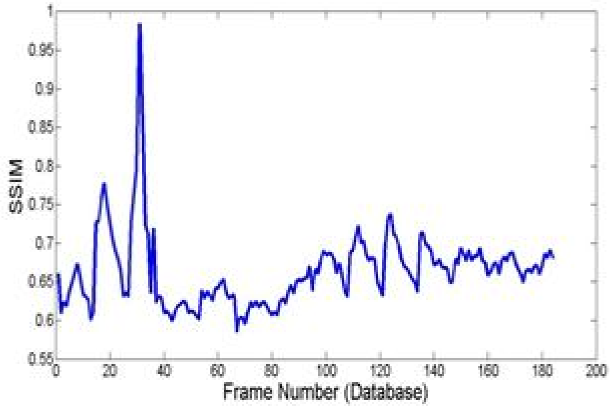


Fig. 8. SSIM Value for a reduced database (Leaving 5 frames from the database)

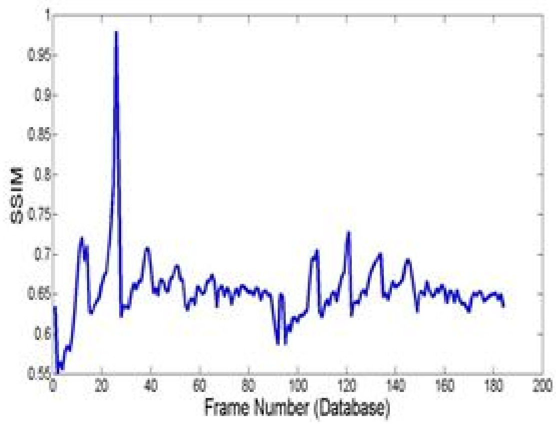


Fig. 9. SSIM Value for a reduced database (Leaving 10 frames from the database)

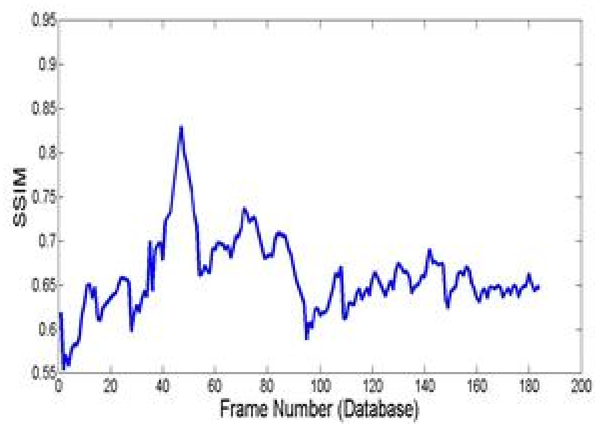


Fig.10. SSIM Value for a reduced database (Leaving 25 frames from the database)

identificationbased on structural similarity index. The method gavegood results for cases with large number of training samplesfor a pre-determined user and could be a good tool for apose authentication system. Current work focuses on applyingpose identification algorithm for a vehicle driver distractionsystem. While it was seen that a reduced database resulted in a reduced confidence in the obtained result, an authenticationproblem would start with possibility to have a large database, under which the proposed method of pose identification givesextremely accurate output.

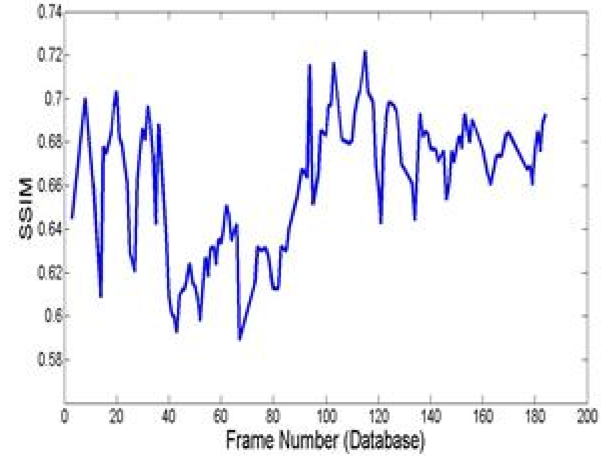


Fig.11. SSIM Value for a reduced database (Leaving 50 frames from the database)

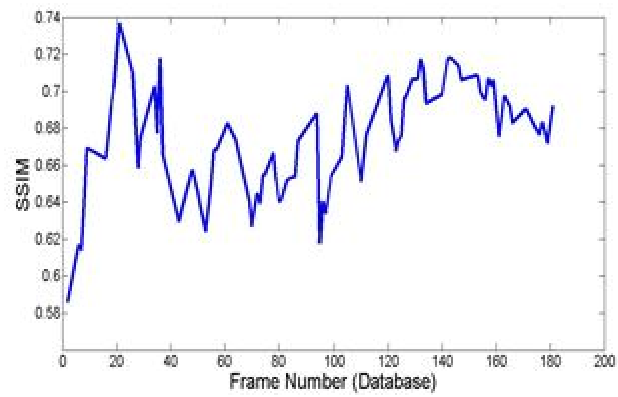


Fig.12. SSIM Value for a reduced database (Leaving 100 frames from the database)

(a) Frame Number (b) Detected Pose from the Reduced Database

Fig.13. Detected Pose in a Reduced Database (Leaving 10 frames from the database)

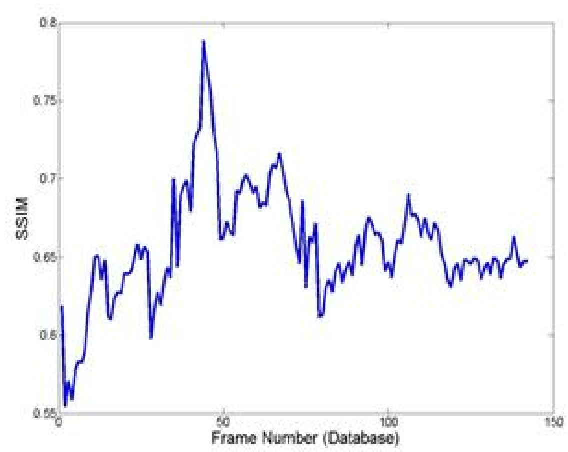


Fig.14. SSIM Values for the frame shown in Fig.13

DOI: 03.LSCS.2013.2.81

ACEEE

References

[1] E. Murphy-Chutorian and M. Trivedi, "Head pose estimation in computervision: A survey," IEEE Transactions on Pattern Analysis andMachine Intelligence,, vol. 31, no. 4, pp. 607–626, 2009.

[2] S. Niyogi and W. Freeman, "Example-based head tracking," in Proceedingsof the Second International Conference on Automatic Face andGesture Recognition,1996.,. IEEE, 1996, pp. 374–378.

[3] D. Beymer, "Face recognition under varying pose," in IEEE ComputerSociety Conference on Computer Vision and Pattern Recognition, 1994.Proceedings CVPR'94., 1994. IEEE, 1994, pp. 756–761.

[4] J. Sherrah, S. Gong, and E. Ong, "Face distributions in similarity spaceunder varying head pose," Image and Vision Computing, vol. 19, no. 12,pp. 807–819, 2001.

[5] G. Goudelis, A. Tefas, and I. Pitas, "Automated facial pose extractionfrom video sequences based on mutual information," Circuits andSystems for Video Technology, IEEE Transactions on, vol. 18, no. 3,pp. 418–424, 2008.

[6] T. Pappas, R. Safranek, and J. Chen, "Perceptual criteria for imagequality evaluation," Handbook of image and video processing, pp. 669–684, 2000.

[7] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment:From error visibility to structural similarity," Image Processing,IEEE Transactions on, vol. 13, no. 4, pp. 600–612, 2004.

[8] Z. Wang and A. Bovik, "A universal image quality index," SignalProcessing Letters, IEEE, vol. 9, no. 3, pp. 81–84, 2002.

[9] "Head pose angles," 2012. [Online]. Available: http://msdn.microsoft.com/en-us/library/jj130970.aspx

[10] N. Gourier, D. Hall, and J. Crowley, "Estimating face orientation fromrobust detection of salient facial structures," in FG Net Workshop onVisual Observation of Deictic Gestures, 2004, pp. 1–9.

DOI: 03.LSCS.2013.2.81