iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 3 Issue: VI Month of publication: June 2015

DOI:

www.ijraset.com Volume 3 Issue VI, June 2015
IC Value: 13.98 ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Active Shape Model Based Pose Estimation Using Hausdorff Matching

Arjun B S1, Dr. Vivek Maik[2]
[1]PG Scholar, Department Of Electronics and Communication Engineering
[2]Head Of the Department, Department Of Electronics and Communication Engineering

The Oxford College Of Engineering, Bangalore, IndiaAbstract— paper presents a novel algorithm for automatic locating, extracting and recognizing human pose. The proposed algorithm consists of four stages; (i) training of typical poses in 2D and 3D configurations, (ii) Hausdorff matching for locating the pose, (iii) silhouette representation of the located pose using deformable templates, and (iv) joint extraction from the silhouette. In order to reduce the computational overhead we used a priori obtained training set, Hausdorff matching with PCA, and a heuristic approach. The algorithm was tested on typical human poses such as; standing, sitting, walking, crouching, and stretching, and Experimental results show the efficiency and accuracy of the proposed algorithm.
Index Terms— Human pose recognition, Human Pose Representation, Heuristic Approach.

I. INTRODUCTION

Physical structure of a human body, which is a combination of bones, joints, and exterior skin, boasts a high level of intricacy and complexity. Human body is also capable of making a wide gamut of different poses, enabled by both innate and acquired flexibility and mobility. Due to the countless number of possible poses, human pose recognition of a 2D image has thus became essential in many applications, such as surveillance, man-machine interaction, medical rehabilitation facilities, content-base retrieval, to name a few. In this paper, we propose a method that identifies, localizes, extracts, and recognizes a human pose from a given image. To the end, we first define the human pose as a combination of meaningful points which are invariant under some transformations. More specifically assign fourteen landmark points at the most important locations, which include joints, such as elbow and knees, and the end points like heads, hand, and feet. Such points are estimated in the model image and consequently used to estimate a 3D body configuration. After defining the pose, we localize the pose region from the input image by applying the Hausdorff distance matching. Because of heavy computation of Hausdorff matching, we used principal component analysis to reduce the dimensions of the pose in advance. In addition, we adopt a model-based approach to extract the silhouette of the human to exploit the robustness of silhouette-based image descriptors. For recognition the action of a pose, we first extract the most probable joint location from the silhouette by using the model-based approach. This method reveals the joint locations that are most useful for pose recognition. Genetic algorithm is then applied to estimate the rest of the necessary joint points that have been omitted in the previous step. The proposed genetic algorithm can accurately estimate the positions of the joints such as elbows, knees, and hips because it minimizes the error between the estimated and the real positions of the joints. Finally, we match the estimated points from the image with the skeletonized training set to recognize the pose. The set of joints obtained by proposed approach is used to reconstruct a 2D pose, and it is refined until it matches with the closest training pose in the sense of both local and global measures.

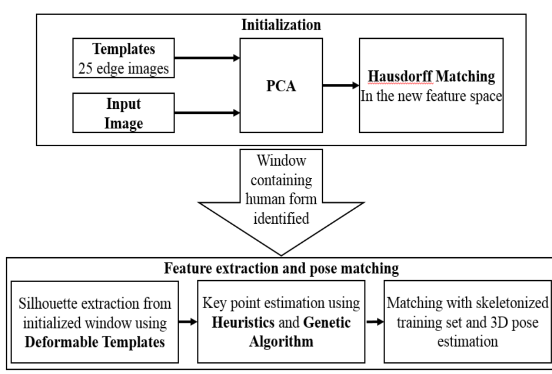


Fig.1. Illustrates the block diagram of the proposed algorithm.

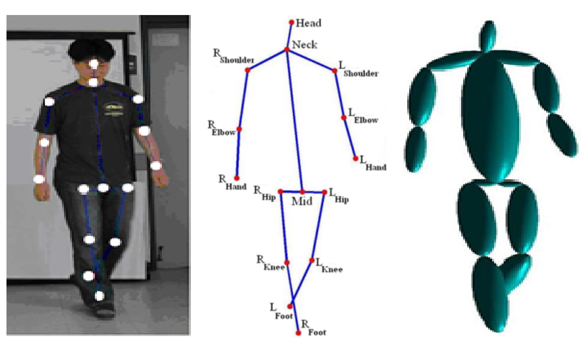
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

II. EXISTING STATE-OF-THE-ART METHODS

Various human pose estimation algorithms have been proposed in the literature with application such as tracking and recognition people in images [1]. Human tracking-based methods have been extended to activity recognition of people in [2]. These methods work well for sequences where robust background subtraction and segmentation algorithms are available. Pose estimation from a video sequence has used spatio-temporal templates and silhouette-based regressive methods [3], [4]. The segmentation and classification algorithms tend to provide reasonable results but still insufficient for the pose application. Recent works on pose estimation from static images using deformable templates and shape contexts have been proposed [5]. Appearance-based methods differ from the proposed method in that the features are related to the whole human appearance and not associated to specific body component. Similar global feature-based methods have been proposed using wavelet features and support vector machines (SVM) classifier [6-8]. Since these algorithms extract the entire human body without locating individual body parts, they are more suitable for applications such as pedestrian detection (walking and standing). Part-based approach used for detecting and estimating people by various body parts (torso, limbs, and head) has been proposed [9]. Another approach for pose estimation is to first locate the body joints and then compute the pose using inverse kinematics [10]. For 3D pose estimation, matching-based methods can been used, which consists of matching features of an input image to a large set of labeled images containing all the possible body poses. In these methods different image features have been used including edges, silhouettes outlines and shape context distributors [11] [12]. In the proposed approach we have combined the matching-based part-based methods into an integrated hybrid framework. As a result we can overcome the drawbacks of both algorithms which include ambiguity in joint estimation and insensitive nature of silhouette descriptors to pose attribute features.

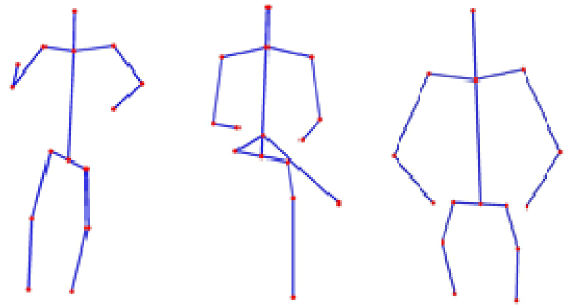
III. HUMAN POSE REPRESENTATION

We make a point distribution model (PDM) that represents the articulated structure of the human body. Using the PDM a pose of an object can be represented by a set of n points, which may be in two or three dimensions. We first obtain a training set of sample points from the image by manually marking the 2D image positions of 14 key points located at body joints (wrists, elbows, shoulders, hips, knees, ankles, head and waist). The selected training points are then used to construct a 2D skeletal representation of the human body pose. The initial training was carried out for sitting, standing, walking, crouching and certain stretching poses for different subjects. In the test images, the ranges of depth of the human body is relatively small compared to the distance between the camera and the person; therefore we adopt a projective transformation for generating 3D models from 2D skeletons. For 3D representation we compare the given articulated pose with the corresponding reference pose of true scale. The projection of the distance between two joints of a test image with respect to that of the reference image will give us the depth for the 3D representation. Potential ambiguity in the depth information for 3D pose configuration is overcome to some extent using kinematics constraints of the human body. After extraction of 2D and 3D training configurations we perform similarity transformation to scale, align and translate them with respect to the reference co-ordinate axes. The transformation parameters are chosen from the mean pose for a particular class of pose (standing or sitting etc).

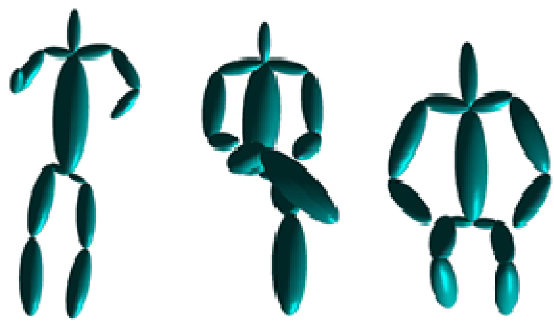


(a) (b) (c)

International Journal for Research in Applied Science & Engineering Technology (IJRASET)



(d)



(e)

FIG. 2: HUMAN POSE REPRESENTATION AND TRAINING :(A) INPUT IMAGE WITH MARKED JOINTS, (B) 2D SKELETAL REPRESENTATIONOF (A), (C) 3D ELLIPSOIDAL MODEL OF THE POSE, (D)2D TRAINING EXAMPLES FOR STANDING, SITTING, AND CROUCHING POSES, AND (E) THE CORRESPONDING 3D MODELS.

IV. LOCATION OF POSE IN STATIC IMAGES

The proposed algorithm begins with locating the pose in a given image. From the located pose we extract the silhouette representing the boundary of the human shape. The processes of identifying the location and extracting the silhouette simplify the feature extraction problem, which would otherwise be very tedious and cumbersome. However the proposed automatic algorithm can accurately locate and extract the silhouette without the need of any user intervention.

A. Human Pose Localization

Given a training set (with labeled key points) and a test image, we can initialize the pose in the image. In this approach we have used Hausdorff distance-based template matching [13] for locating poses in the image. For each category of poses (standing, sitting, crouching and stretching) a particular single reference template was generated as input for Hausdorff distance matching. Since this step is only used to locate the pose in the object the input templates need not be exactly the same pose pattern to one in the test image. One of the drawbacks of using the Hausdorff distance is the high computational load. For a given test image, locating the pose using several test templates is also time consuming. So in our method we have used PCA to reduce dimension before applying the Hausdorff distance for pose pattern matching. In our approach a pose template is represented by a discrete set of k points for n different poses as

$$P = \{p_1^k, ..., p_n\},\ p_i \in R^2, \qquad (1)$$

Since the hausdorff optimization is carried out on edge templates we obtain the edge contours of both input and a template using canny edge operator as

$$h_{\max} = \sum_{x=1}^{M}\sum_{y=1}^{N}\{\hat{p}_i(x,y) \subseteq \hat{q}(x,y)\}, \qquad (2)$$

where $h_{\max}$ is the maximum hausdorff distance between the subset of input and template edge pixels [3] and $(\hat{p}_i, \hat{q})$ represents

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

the edge map of the template and the image. The subset of pixels was selected for window of size 20×30 pixels. For PCA we calculate the covariance matrix S for both templates and the input image. From the covariance matrix the set of eigenvectors for the templates, $\phi_p = \{\phi_1,...,\phi_n\}$, and the input image ϕ_q are computed. Now we can select the first few eigenvectors corresponding to the largest eigenvalues. If $\phi_p = \{\phi_1,...,\phi_n\}$ and ϕ_q contains the t largest eigenvectors corresponding to 90% of largest eigenvalues. From the new subset of eigenvectors we can generate the feature vectors for the templates p_f and the input image q_f as,

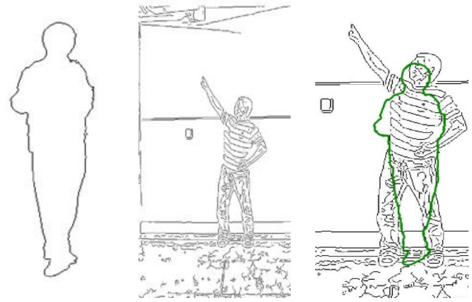


Fig.3. Illustration of Hausdorff matching

$$p_f = (\phi_p)^T * (\hat{p}_i)^T \textit{ and } q_f = (\phi_q)^T * (\hat{q}_i)^T . \qquad (3)$$

The Hausdorff matching is now carried in the new feature space, at the significantly reduced number of dimensions and computation complexity as shown in Fig.3.

B. Pose Silhouette Extraction

Of the many different image descriptors for human pose estimation we have chosen image silhouettes for the proposed system. The main advantage of the silhouette- based image descriptor is that it can be robustly extracted from images and are insensitive to irrelevant surface attributes, such as clothing color and texture. Even though silhouettes are global measures, they still possess a great deal of pose variation making them suitable for pose recognition applications. Several classes of shape model- based techniques for extracting silhouettes have been proposed [14]. In this algorithm we have used active shape model (ASM-based approach to extract the silhouette of the human pose. The proposed ASM is slightly different from the one proposed by Cootes and Taylor [15]. After fitting of a shape model to a test image we have used energy minimization for locating the boundary points instead of sample profile search method. The energy minimization approach, which was originally proposed as active contours or snakes, was used in conjunction with ASM to provide a more localized boundary extraction. The templates used for silhouette extraction using deformable templates were the same as those used for the Hausdorff matching. However with the initial location of the pose we carry out the minimization to extract the silhouette using deformable templates. For a given standing pose we can extract the silhouette using the corresponding template and deformable optimization. However the extracted silhouettes do not contain a sufficient amount of information for recognizing a pose.

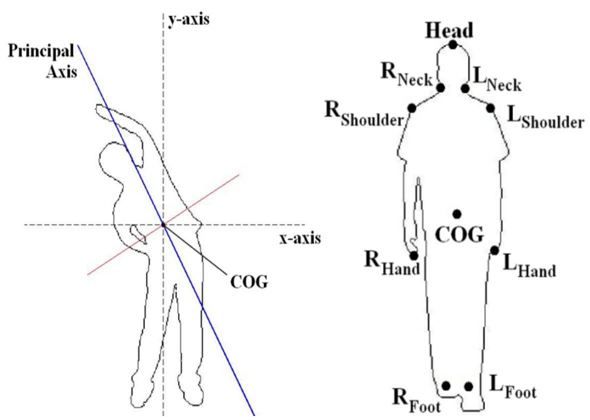
Volume 3 Issue VI, June 2015
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

V. EXTRACTION OF JOINT FEATURES FROM SILHOUETTES

Silhouettes leave several discrete and continuous degrees of freedom invisible or poorly visible. The extracted silhouettes have to be processed for further pose identification and recognition. By extracting silhouettes we can superficially classify the pose into one of the typical classes, such as standing or sitting etc. but we need to define or recognize the action of a pose more specifically. For instance, a standing person can drink coffee or talking on a phone, which can be only, recognized using specific body joint features and training data. In the recognition phase we initially extract the features from the silhouette data using a heuristics approach explained below.

A. Heuristic Approach To Joint Estimation

The extraction of most probable joint location from the extracted silhouette is carried out using a heuristics approach. This method significantly eliminates the pool of potential feature candidates revealing only joint location, which is useful for pose identification. As a prerequisite for joint estimation we use the following derivation. The center of gravity (COG) of the Euclidean distance transformed image is estimated to locate the principal axis of the pose. The Euclidean distance transform of the silhouette of the image gives a skeletal representation that makes estimation of the joints more accurately,



(a) **(b)**

Fig.4 : Heuristic approach to joint estimation ; (a) the principal axes located at the eigenvectors and (b) the silhouette with the estimated joints using the proposed approach

$$x_g = \frac{1}{N}\sum_{i=1}^{N} x_i,\ y_g = \frac{1}{N}\sum_{i=1}^{N} y_i\ . \qquad (4)$$

where (x_g, y_g) represents the COG of the i^{th} pixel of the distance averaging image. For estimating the joints we need to compute the principal axis or axis of symmetry of the respective silhouette. The COG along with the inclination of the body would give the axis of symmetry of a particular pose. In our approach we have used the distance of the i^{th} pixel compute the axis along the COG. The COG and the axes of symmetry now form the basis for extraction of joint from the silhouette image. Feet are identified as the farthest points from the COG, in the lower half of the body, on either side of the PA. The torso is separated from the body by splitting it in the ratio $1/r$ (upper: lower),where r is selected on the basis of symmetry conditions. Intersection of the principal axes of the torso with its silhouette gives the top of the head. An asymmetric pose is the one whose principal axes differ substantially from that of the torso. We have found that for a symmetric pose, $r = 2$ and, for unsymmetrical pose, $r = 2.25$ provides acceptable results for most experiments. For the upper half of the body, the points of the contour that are closest to the principal axis of the torso, on either side are the two neck points. The silhouettes between the left (right) neck and the left (right) foot are divided in to two halves, and the midpoint is the approximated waist. The upper half contains the left (right) arm. For each arm there can be three upper, mid and lower joints. The point almost in parallel and closest to neck on either side is identified as the shoulder or upper point. From that point the lower tip can be found as the farthest in that length. The proposed joint estimation process is illustrated in Fig.4.

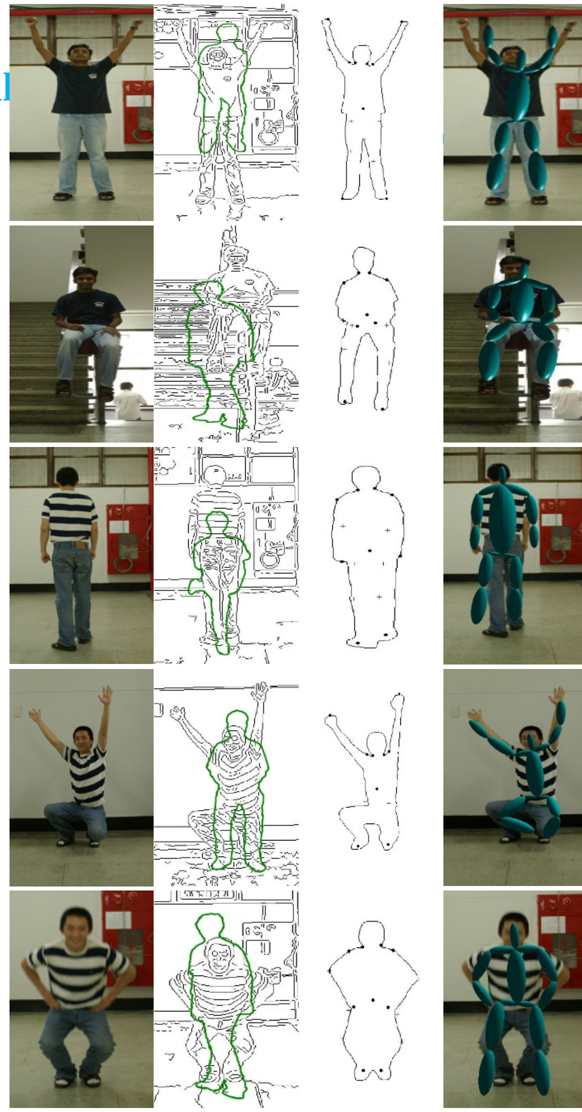
B. Genetic Algorithm For Recovering Missing Joints

Heuristic method can estimate all joint locations with sufficiently high accuracy for the recognition stage. We found that for natural poses we were able to estimate 10 joints, so the remaining joints (elbows, knees and hips) should be estimated using a genetic algorithm which utilizes the estimated joints. For training the genetic algorithm we have used several images in the training set. The genetic algorithm minimizes the error between the estimated and real positions of the joint. A naïve method of doing this would be to calculate the errors for each database image and minimize them individually by varying the coefficient matrices. This process can be conveniently replaced by a simpler approach, using the fitness function:

$$f = \frac{1}{N}\sum d_i + \frac{1}{N-1}\sum (d_i - \overline{x})^2\ , \qquad (5)$$

where d_i represents the distance between the estimated and the real locations of the required joint position, and $\overline{x}$ the mean error of the trained database.

VI. POSE MATCHING RESULTS



(a) (b) (c) (d)

Fig.5 : Experimental results of pose recognition : (a) Input image, (b) Hausdorff matching for initial pose, (c) heuristic joint estimation from the sillouette representation, and (d) 3D pose matching and recognition.

The joints estimated by using the proposed approach reconstruct a 2D pose. This gives an abstract version of the pose as it may have some misalignments and misinterpretations. The reconstructed pose is aligned, scaled and translated with respect to the training set. The reconstructed pose is again refined to match with the closest training pose. The reconstructed pose is matched both locally (individual key points) and globally (entire pose). We used a training database of 80 frontal poses to infer the 3D pose from the key points extracted from the silhouettes. From the matched 2D pose the corresponding 3D pose model is generated. For matching we have used the length and angle extracted from the joints. The measure of similarity between joints can be estimated using standard deviations of the lengths and angles in the input image from those of the training set. For poses which do not exists in the database a new pose is generated directly from the estimated features. Fig. 5 represents the results of the proposed algorithm for various poses including standing, walking, sitting, and stretching. The poses in the first three rows have a corresponding match in the training set; hence the pose is matched with the training set. The remaining poses are the ones which does not have a corresponding match in training database hence their pose is generated from the joint features. The Hausdorff matching results are shown the second column. The Hausdorff templates are chosen automatically from the test templates. The templates which require minimum computation time are chosen for a particular image. The computation time for the Hausdorff matching with PCA and 25 test templates was found to be 10 sec/image. The estimated joints using the proposed heuristic approach are indicated using dots

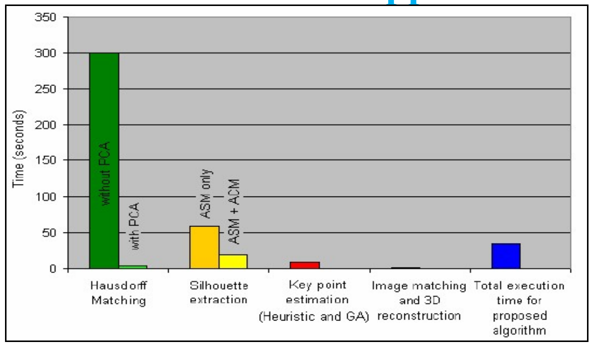


Fig.6 : Graphical representation of the computation time for the proposed algorithm

whereas those estimated by using the genetic approach are represented by arrows. The final reconstructed pose from the estimated joints is shown in the last column. The proposed algorithm was found to perform well for large class of poses with differing variations.

VII. CONCLUSION

The main contribution of the proposed algorithm is to automatically identify, locate, extract, and recognize pose from static images. The pose was modeled using the reduced number of joints for acceptable results. The use of a priori built training set along with the Hausdorff distance using PCA and a heuristic approach for joint estimation significantly reduce the computation complexity, and improve the speed of processing. Future work will focus on extending the view-point invariant algorithms and able to recognize multiple poses in one image. Also another useful contribution would be to enhance the heuristic approach to overcome occlusion within the pose and extract all the joints from the given silhouette. Experimental results suggest that the proposed method performs well for a given frontal pose.

REFERENCES

[1] M. Lee and I. Cohen, “A model-based approach for estimating human 3D poses in static images,” IEEE Trans. Pattern Analysis, Machine Intelligence, vol. 28, no. 6, pp. 305-317, June 2006.
[2] M. Harville and D. Li, “Fast, integrated person tracking and activity recognition with plan-view templates from a single stereo camera,” Proc. IEEE Conf. Computer Vision, Pattern Recognition, vol.2, pp. 398-405, July 2004.
[3] A. Elgammal and C. Lee, “Inferring 3D body pose from silhouettes using activity manifold learning,” Proc. IEEE Conf. Computer Vision, Pattern Recognition, vol. 2, pp. 681-688, July 2004.
[4] A. Agarwal and B. Triggs, “3D human pose from silhouettes by relevance vector regression,” Proc. IEEE Conf. Computer Vision, Pattern Recognition, vol. 2, pp. 882-888, July 2004.
[5] G. Mori and J. Malik, “Recovering 3D human body configurations using shape contexts,” Proc. IEEE Trans. Pattern Analysis, Machine Intelligence, pp. 1052-1062, June 2006.
[6] L. Bogoni and M. Hansen, “Pattern Selective Color Image Fusion,” Int. Journal, Pattern Recognition, vol.34, pp. 1515-1526, 2001.
[7] O. Chappelle, P. Haffner and V. Vapnik, “Support vector machines for histogram based image classification,” IEEE Trans. Neural Networks, vol. 10, no. 5, pp. 1055-1064, September 1999.
[8] J. B. Arie, Z. Wang, P. Pandit, S. Rajaram, “Human activity recognition using multi dimensional indexing,” IEEE Trans. Pattern Analysis, Machine Intelligence, vol. 24, no. 9, pp. 1091-2005, August 2002.
[9] A. Mittal, L. Zhao and L. Davis, “Human body pose estimation using silhouette shape analysis,” Proc. IEEE. Conf. Advanced Video and Signal based Surveillance, pp. 263-270, July 2003.
[10] K. Grumman, G. Shakhnarovich, and T. Darrell, “Inferring 3D structures with statistical image based model,” Proc. IEEE Conf. Computer Vision, vol. 1, pp. 941-647, June 2003.
[11] J. Shin, H. Ki, V. Maik, J. Kang, J. Jung, and J. Paik, “Evolutionary algorithm-based local structure modeling for improved active shape model,” Proc. EvoWorkshops, vol. 3005, pp. 359-368, April 2004.
[12] A. Koschan, S. Kang, J. Paik, B. Abidi, and M. Abidi, “Color active shape models for tracking non-rigid objects,” Pattern Recognition Letters, vol. 24, no. 11, pp. 1751-1765, July 2003.
[13] Y. Gao, “Efficiently comparing face images using a modified Hausdorff distance,” IEE Proc. Vision, Image and Signal Processing, vol. 150, no. 6, pp. 346-350, December 2003.
[14] D. Kim, V. Maik, D. Lee, J. Shin, and J. Paik, "Active shape model-based object tracking in panoramic video, " Proc. ICCS2006, LNCS, vol. 3994, pp. 922-929,

Volume 3 Issue VI, June 2015
ISSN: 2321-9653

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

May 2006.

T. Cootes, D. Taylor, and C. Abraham, “Active shape models- their training and applications,” Computer Vision and Image Understanding, vol. 61, no. 1, pp 38-59, January 1995.