

Detection and Pose Estimation of People in Images

Mousa Mojarrad, Amir Masoud Rahmani, and Mehrab Mohebi

Abstract—Detection, feature extraction and pose estimation of people in images and video is made challenging by the variability of human appearance, the complexity of natural scenes and the high dimensionality of articulated body models and also the important field in Image, Signal and Vision Computing in recent years. In this paper, four types of people in 2D dimension image will be tested and proposed. The system will extract the size and the advantage of them (such as: tall fat, short fat, tall thin and short thin) from image. Fat and thin, according to their result from the human body that has been extract from image, will be obtained. Also the system extract every size of human body such as length, width and shown them in output.

Keywords—Analysis of Image Processing, Canny Edge Detection, Human Body Recognition, Measurement, Pose Estimation, 2D Human Dimension.

I. INTRODUCTION

THE goal of the pose estimation and feature extraction module are to take a large amount of image data and obtain information necessary to identify or distinguish the human. As in the case of feature extraction, knowledge of the image domain may be used to affect the feature grouping process. The analysis and use of visual information is a first order task for AI researchers. Due to the architecture of classical computers and to the computational complexity of state of the art algorithms, it is required to find better ways to store, process and retrieve information for image processing. Digital image processing is the science of processing imagery a visual media via a digital computer. As a visual media, while a great deal can be learned by discussing the theory of image processing, the greatest understanding of image processing naturally comes from viewing processed digital images. The ability to detect people in images is key to a number of important applications ranging from surveillance, robotics, and intelligent vehicles to advanced user interfaces [1]. Recognition of human and extraction their features in images are a major issue in detecting objects, since the positions of

Mousa Mojarrad is with the Islamic Azad University Firoozabad branch, Iran. He is now member of faculty of Islamic Azad University Firoozabad branch, Iran. (Phone: +987126229701-3; fax: +987126224402; e-mail: m_mojarrad@iauf.ac.ir).

Amir Masoud Rahmani is with the University of Sciences and Researches Tehran branch, Iran. Now he is head of department computer engineering of University of Sciences and Researches Tehran branch, Iran (e-mail: rahmani@sr.iau.ac.ir).

Mehrab Mohebi is now member of faculty of Islamic Azad University Firoozabad branch, Iran.

Human in the picture can be different, such as sitting or standing, opposite or back to the image, clothed or unclothed. Also one of the important challenges in finding human segmentation in the image is the issue of modeling human shape in the image [1]. Parameter estimation is important and is usually implemented for solving some medicine Problems as analysis of the biomedicine images, in military industry; help safeguard public spaces such as train stations, shopping malls, street corners, medical, multimedia Applications and in view of mounting concerns about public safety, etc.

II. PREVIOUS WORKS

There are a few invariant feature extraction methods known, nowadays. They are used in various parts of signal and image processing research fields. One of them is computer vision field, where the scientists try to solve the problems introduced with interaction of automated system with reality using the information represented by image and transformation of this information to object, modeling the aspects of reality useful for other image processing steps. These automated systems can be used for change detection [3], invariant image recognition [4] and for image parameter estimation [5]. There are many well-known and possible methods of invariant feature extraction like Zernike moments, Radon transform, or some new methods like Trace transform [3], usable in parameter estimation tasks, too.

Recent approaches to person detection exploit articulated body models in which the body is viewed as a kinematic tree in 2D [6], 2.5D [7], or 3D [8] leading to a parametric state-space representation of roughly 25–35 dimensions. The high dimensionality of the resulting state-space has motivated the development of specialized stochastic search algorithms that either exploit the highly redundant dynamics of typical human motions [9], or use hierarchical sampling schemes to exploit the tree-structured nature of the model [10]. Hierarchical body models lead to “top-down” search algorithms that make it difficult to incorporate “bottom-up” information about salient body parts available. As a result, few, if any, of the above methods deal with the problem of automatic initialization of the body model. Finally, the fully coupled kinematic model results in a computationally challenging search problem because the search space cannot be naturally decomposed. Variations on this type of model have been recently applied by Burl, Weber and Perona [11] and Ioffe and Forsyth [12, 13].

The main benefits are that it supports inference algorithms where the computational cost is linear rather than exponential in the number of body parts, it allows elegant treatment of occlusion, and it permits automatic initialization based on individually unreliable low-level body part detectors [14]. The work described here, like the previous work above, exploits this notion of flexible “spring”-like constraints [15] defined over individually modeled body parts [12,13], though we extend the approach to locate the parts in 2-space rather than the 3-dimensional image plane. The body is treated as a graphical model [16], where each segment in the graph corresponds to an independently parameterized body part. The spatial constraints between body parts are defined as directed edges in the graph.

Hochberg and Brooks [17] describe the case of a 19-month-old human baby who had had no previous exposure to any kinds of two-dimensional images, yet was immediately able to recognize ordinary line drawings of known objects. It is true that there has been some research on the bottom-up derivation of depth directly from line drawings or the edges detected in a single image [18], including previous research by the author [19].

Several methods have been proposed for identifying people in color/grayscale images. Some examples include the direct use of wavelet features with support vector machines [20], coarse to fine edge template matching [21], motion/intensity AdaBoosted classifiers [22], and the size/shape of image differencing regions [23]. Several other related methods using color, texture, and stereo have also been proposed.

Most of the approaches can be distinguished according to the features used. One popular approach is spin images Frome, Huber, Kolluri, Bulow, and Malik [24]. Other types of 3D features include local tensors Mian, Bennamoun, and Owens [25], shape maps [26], and multi-scale features [27]. Osada, Funkhouser, and Dobkin [28] proposed a 3D object recognition technique based on shape distributions. Whereas this approach requires a complete view of the objects, our method can deal with 2D seen objects.

A feature-based recognition approach recently developed by Lowe [29] consists of four stages: feature detection, extraction of feature correspondences, poses parameter estimation, and verification. Features are computed over multiple scales, at positions that are extrema of a difference-of-Gaussian function. The constellation model [30] also relies on matching image parts, but typically uses on the order of 5 features. One drawback of the constellation model is the high number of training samples required, although recent work by Fei-Fei, Fergus and Perona [31] proved that learning can be efficiently achieved with few examples. Another disadvantage of the constellation model lies in the large computation time required in order to learn feature conjunctions, limiting it to use of a relatively small number of parts for each object category.

III. OBJECT DETECTION METHODS

Many remote sensing applications such as image registration, image segmentation, object detection and recognition, etc, use edge detection as a preprocessing stage for feature extraction. There many ways to perform edge detection. The methods based on the first Derivative such as Roberts, Prewit and Sobel were studied. Also, an implementation of the Canny edge detector is proposed. There are two strategies and solutions for detection of objects [32]: A) Searching based on the model characteristics using a comparison between the object model and the image. B) Summing-up of image features in larger groups.

The top-down approaches are used in most activities in images. Top-down approaches decrease the area searching and also time searching, but increase the complication. In This suggested method have been helped the top-down approach for segmentation and recognition human body and also extracting the Human Body features in image.

IV. SEMI-AUTOMATIC FEATURE EXTRACTION

Automatic feature extraction is an interactive process between an operator and one or more computer algorithms. To initiate the process, the operator interprets the image and decides which features are to be measured and which algorithms are to be used for this task. If required the operator also may tune some of the algorithm’s parameters and select an object model for the current feature. Semi-automatic feature extraction algorithms have been developed for measuring primitive features such as points, lines and regions, but also for more complex, often parameterized, objects.

V. IMAGE ENHANCEMENT

One of the important operations in field of image processing is an image enhancement [33].

This method can be done in area field or frequency field. Some algorithm’s enhancement operates freely on all the image’s pixels that have been populated image’s pixels operation. Another algorithm operates on image’s pixels with pixels neighbors that are famous image’s area operation. In this paper, has been used this method for images enhancement.

VI. RESIZING IMAGE

Images resize done on entire picture in this project .the aim of this Image resizing in this project, are increasing the image overhead. On of the resizing Image method, use the matrix operation [33]. In this method, has been used the window that change the original image into resizing image. Then this window move from beginning to end. Resolutions Average dots that window cover them, used as a dot in the center of window. This subject, as shown in fig.1.

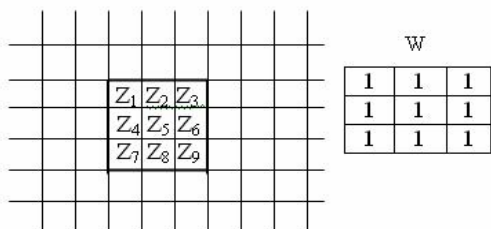


Fig. 1 use window 3× 3 for image resize

After operating window on the image, the z_5 dot resolution, as shown in equation.1.

$$z_5 = \frac{1}{9} \sum_{i=1}^9 w_i z_i \tag{1}$$

The size of the image has been changed by using this method to the 346×461 images. This was done to reduce the burden of computation.

VII. FRAMEWORK OF THE SUGGESTED METHOD

We proposed framework for estimating 2D human body pose which allows us to extract measures for evaluating human body. The typical framework for the suggested method system is shown in Fig 2.

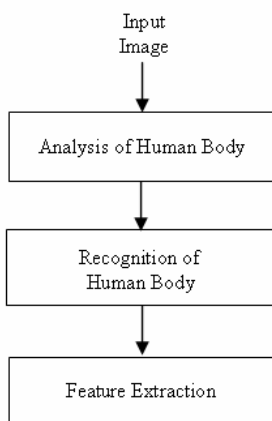


Fig. 2 the Suggested Method for feature Extraction of the Human Body Composition in Images with segmentation

From the input digital image, an appropriate set of features are extracted. In the first stage, the image is received from input, the image edges are extracting by using the canny method and the edge points will be found. The result of this stag is a figure that contains edge points from input image. In the connection of edges stage, by operating on edge points that has been came from previous stage, the edge points turned into edges list. The instruction is, that by searching the first edge point such as (x_1, y_1) ; this point is choose as a first list member. Then by searching eight points around this point (fig.3) every point from this neighbor points that was edge point, will be chosen as a second member of a list. After that from the new neighbor points or the last element of a list, the

new edge point will be searched and also added to the list. This work will be continued till the last point of a list does not have a neighbor edge point. The whole dots that have been gathered are signed for a second list, and then the new list of edge point will be searched. The outputs of this stage are a list of edges.

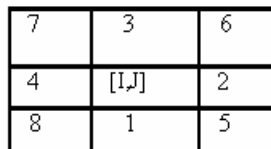


Fig. 3 eight neighbor points around the center point

In the stage of conversion the edges to lines, the connected edge points in past stage, turned into straight lines. In this stage, turning into the edge points to lines are done by the help of up to down, breaking in two lines method. In this method, as it shown in fig.4. First, a line is made between the first point and last point of a list. Then if the maximum normal error (MNE) increased the maximum range, this line will be divided into two lines. As the flowing that among the points, the point that has the most distance from the mentioned line (point C) will be chosen and (AB) line will be divided into (AC, CD)lines. This algorithm will be done on the all edge points of lists and the collection of the lines will be gained.

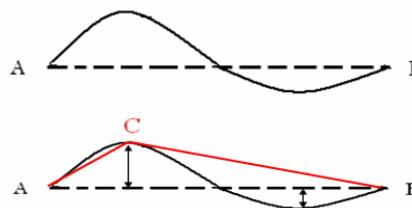


Fig. 4 Conversion of edges to lines by up to down, breaking in two lines method

In the stage of classifying line's slope, line's slope will be gained. After gaining line's slope, they will be classified. Since two parallel lines have the same slope, in this stage the aim is classifying the lines that are based on slope. It is clear if two lines have a same slope, definitely they will be put in the same part. The relation between line's slope and angles with horizontally axis, are shown in Equation.2.

$$\theta = \arctan(m) \tag{2}$$

θ Is a line angle with horizontally axis and M is a line slope. If two lines with a bit error are parallel, they are putt in the same part. For this, twelve sectors are supposed, as are shown in fig.5. These sectors cover each other. In the other hand a line can be a member of two sectors. Finally twelve classes will be gained from the lines.

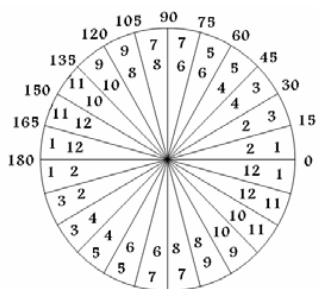


Fig. 5 each line that has a degree with horizontally axis; can be a member of two sectors

After classifying the lines according to their slopes, it turns to; finding the possible built segments by these lines. In the rectangle two lengths are parallel with each other and they are symmetric with the passed axes from the middle of rectangle's width [12]. According to this reason if there are a few parallel or symmetric lines around a hypothesis axis, they can be supposed as a rectangle. For this, all the lists will be tested one by one in order to finding the rectangles.

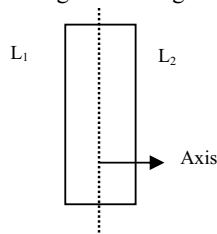


Fig. 6 in the rectangle two lengths are parallel and symmetric with the passed axis of them

In the stage of recognition of human body segments, the human body assumed in standing position that consists of five segments (including: torso, two hands and two legs). Two hands and two legs are connected to the torso in a special position. There are a few attribute in these segments that is useful for finding human body segments in the figures. A) A torso has a basic role for finding human body segments. Because, four other segments (two hands and two legs) are specially connected to the torso. B), as it has been shown in fig.7. The connection points of two hands and two legs are symmetric with a vertically axis that passed from the middle of body. C) For each hand: width and length of hands, and for each leg: width and length of legs can be assumed that they are equal two by two.

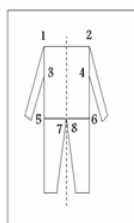


Fig. 7 the shape of five segments of human body (including: torso, two hands and two legs)

With these attributes, segment's position with each other, their area, their width and length, help the algorithm for find a human in picture. Finally Human Body Composition Feature, Extracted from the previous stage.

In the stage of feature extraction, the entire feature extract from image. The measure of segments was measured by the results of previous stages and then, the measures convert by toolbox conversation of Matlab software.

Also, the area of the pictures measure was measured through the segment that obtained in the recognition of human body being in pictures.

VIII. EXPERIMENTS

The Results of executing the proposed method on a colorful images and features extraction from image, as shown in fig.8. The Result of feature extraction of the human body composition, as shown in Table 1, 2. The graph of relationship between height and area are shown in fig.9.

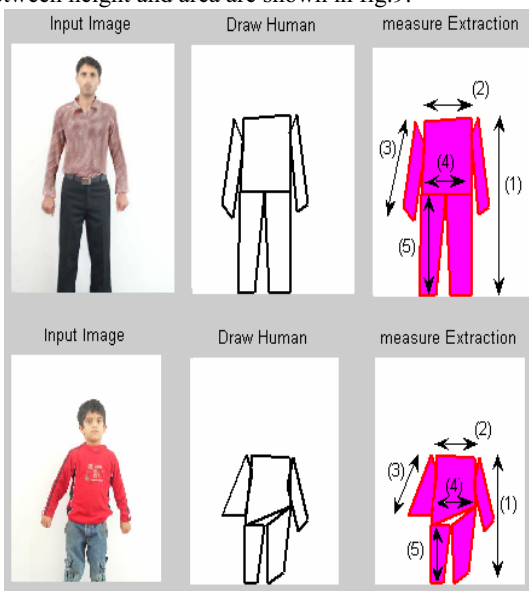


Fig. 8 Results of executing the proposed method on a colorful image and features extraction

TABLE 1 RESULTS OF FEATURE EXTRACTION OF THE HUMAN BODY COMPOSITION FOR FIRST HUMAN.

measure	(1)	(2)	(3)	(4)	(5)
width	-	46cm	-	35cm	-
height	150cm	-	64cm	-	101cm

TABLE II RESULTS OF FEATURE EXTRACTION OF THE HUMAN BODY COMPOSITION FOR SECOND HUMAN.

measure	(1)	(2)	(3)	(4)	(5)
width	-	24cm	-	17cm	-
height	71cm	-	36cm	-	40cm

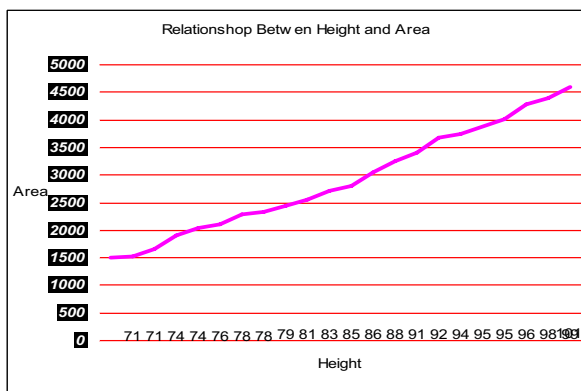
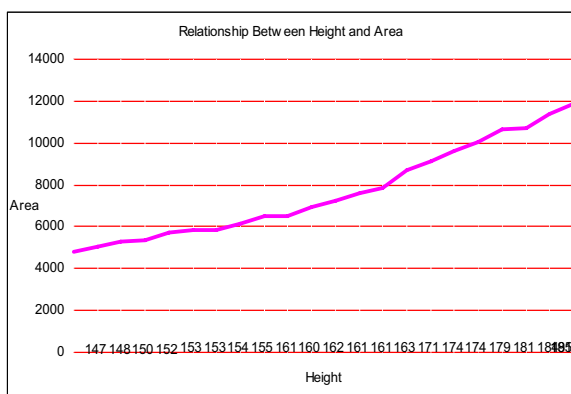


Fig. 9 Graph of Relationship between Height and Area

IX. CONCLUSION

In previous works, only the human body has been recognized in image and also a few measures extracted.

But in this work, has been use Cany edge detection and all the measure extract from image. Furthermore, the advantage of this algorithm are that can obtain all measure and also obtain the area of human body from image that it use for recognize the fat and thin of human body from image, and the classify them.

Meanwhile the designed program can work on all of the human sizes. This algorithm was applied to a number of colorful images containing human body and that 96 percent of obtained results were satisfactory.

REFERENCES

- [1] D.M. Gavrilu, "The Visual Analysis of Human Movement: A Survey," Computer Vision and Image Understanding, vol. 73, no. 1, pp. 82-98, 1999.
- [2] Brett Allen. And Brian Curless. And Zoran Popovi. And Aaron Hertzmann.2006." Learning a correlated model of identity and pose-dependent body shape variation for real-time synthesis," Eurographics/ ACM SIGGRAPH Symposium on Computer Animation 2006.
- [3] KADYROV, M. PETROU: "The Trace Transform and Its Applications", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.23, No.8, pp. 811-828, August 2001
- [4] N. FEDOTOV, L. SHULGA: "New Theory of Pattern Recognition on the Basis of Stochastic Geometry", in Proc. WSCG 2000, the 8-th International Conference in Central Europe on Computer Graphics, Visualisation and Digital Media 2000, Plzeň, Czech Republic, Feb. 2000.
- [5] G. WOLBERG, S. ZOKAI: "Robust Image Registration using Log-polar Transform", wwwcs. engr.ccnyc.edu/~wolberg/pub/icip00.pdf (last visit 20.7.2003).
- [6] S. Ju, M. Black and Y. Yacoob. Cardboard people: A parameterized model of articulated motion. Int. Conf. on Automatic Face and Gesture Recognition, pp. 38-44, 1996.
- [7] Y. Wu, G. Hua and T. Yu, Tracking articulated body by dynamic Markov network, ICCV, pp. 1094-1101, 2003.
- [8] C. Sminchisescu and B. Triggs. Covariance scaled sampling for monocular 3D body tracking, CVPR, vol. 1 pp. 447-454, 2001.
- [9] H. Sidenbladh, M. Black and D. Fleet. Stochastic tracking of 3D human figures using 2D image motion, ECCV, vol. 2, pp. 702-718, 2000.
- [10] J. MacCormick and M. Isard. Partitioned sampling, articulated objects, and interface-quality hand tracking. ECCV (2), pp. 3-19, 2000.
- [11] M. Burl, M. Weber and P. Perona . A probabilistic approach to object recognition using local photometry and global geometry, ECCV, pp. 628-641, 1998.
- [12] S. Ioffe and D. Forsyth. Probabilistic methods for finding people, IJCV 43(1):45-68, 2001.
- [13] D. Ramanan and D. Forsyth. Finding and tracking people from the bottom up, CVPR, Vol. II, pp. 467-716, 2003.
- [14] S. Yu, R. Gross, and J. Shi. Object segmentation by graph partitioning Concurrent object recognition and segmentation by graph partitioning, Advances in Neural Info. Proc. Sys. 15, pp. 1407-1414, 2003.
- [15] P. Felzenszwalb and D. Huttenlocher. Efficient matching of pictorial structures, CVPR, Vol. 2, pp. 66-73, 2000.
- [16] M. Jordan, T. Sejnowski and T. Poggio. Graphical models: Foundations of neural computation, MIT Press, 2001.
- [17] Hochberg, Julian E. and Virginia Brooks, "Pictorial recognition as an unlearned ability: A study of one child's performance," American Journal of Psychology, 75 (1962), 624-628.
- [18] Barnard, Stephen T., "Interpreting perspective images," Artificial Intelligence, 21 (1983), 435-462.
- [19] Lowe, David G., and Thomas O. Binford, "The recovery of three-dimensional structure from image curves," IEEE Trans. on Pattern Analysis and Machine Intelligence, 7, 3 (May 1985), 320-326.
- [20] M. Oren, C. Papageorgiour, P. Sinha, E. Osuma, and T. Poggio." Pedestrian detection using wavelet templates". In Proc. Comp. Vis. and Pattern Rec., pages 193-199. IEEE, 1997.
- [21] D. Gavrilu. Pedestrian detection from a moving vehicle. In Proc. European Conf. Comp. Vis., pages 37-49, 2000.
- [22] P. Viola, M. Jones, and D. Snow. "Detecting pedestrians using patterns of motion and appearance". In Proc. Int. Conf. Comp. Vis., pages 734-741, 2003.
- [23] Lipton, H. Fujiyoshi, and R. Patil. "Moving target classification and tracking from real-time video". In Proc. Wkshp. Applications of Comp. Vis., 1998.
- [24] Frome, D. Huber, R. Kolluri, T. Bulow, and J. Malik. "Recognizing objects in range data using regional point descriptors". In Proc. of the Europ. Conf. on Computer Vision (ECCV), 2004.
- [25] Ajmal S. Mian, Mohammed Bennamoun, and Robyn A. Owens. "Matching tensors for automatic correspondence and registration". In Proc. of the Europ. Conf. on Computer Vision (ECCV), 2004.
- [26] Z. Wu, Y. Wang, and G. Pan. "3D face recognition using local shape map". In Proc. of IEEE Intern. Conf. on Image Processing, pages 2003-2006, 2004.

- [27]] X. Li and I. Guskov. "Multiscale features for approximate alignment of point-based surfaces". In Symp. on Geometry Processing, pages 217–226, 2005.
- [28] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin. "Matching 3D models with shape distributions. In Shape Modeling International", pages 154– 166, 2001.
- [29] D.G. Lowe: "Distinctive image features from scale-invariant keypoints" accepted paper, Int. J. of Comp. Vision, 2004.
- [30] M. Weber: "Unsupervised Learning of Models for Object Recognition", Ph.D thesis, Department of Computation and Neural Systems, California Institute of Technology, Pasadena, CA, 2000
- [31] L. Fei-Fei, R. Fergus and P. Perona: "A bayesian approach to unsupervised one-shot learning of object categories", Proc. Int. Conf. on Comp. Vision, Nice, France, 2003.
- [32] S. Ioffe and d.a. Forsyth.2001."Probabilistic Methods for Finding People," International Journal of Computer Vision 43(1), 45–68, 2001.
- [33] Gonzales R.C.and Woods R.E, (1992)."Digital Image Processing", USA, Addison-wesley.