Journal of Computer and Communications, **2014, 2, 78-84**
Published Online January 2014 (http://www.scirp.org/journal/jcc)
http://dx.doi.org/10.4236/jcc.2014.22014

Multi-Scale Human Pose Tracking in 2D Monocular Images

Jinglan Tian, Ling Li, Wanquan Liu

Department of Computing, Curtin University, Perth, Australia.
Email: jinglan.tian@student.curtin.edu.au

Received October 2013

ABSTRACT

In this paper we address the problem of tracking human poses in multiple perspective scales in 2D monocular images/videos. In most state-of-the-art 2D tracking approaches, the issue of scale variation is rarely discussed. However in reality, videos often contain human motion with dynamically changed scales. In this paper we propose a tracking framework that can deal with this problem. A scale checking and adjusting algorithm is proposed to automatically adjust the perspective scales during the tracking process. Two metrics are proposed for detecting and adjusting the scale change. One metric is from the height value of the tracked target, which is suitable for some sequences where the tracked target is upright and with no limbs stretching. The other metric employed in this algorithm is more generic, which is invariant to motion types. It is the ratio between the pixel counts of the target silhouette and the detected bounding boxes of the target body. The proposed algorithm is tested on the publicly available datasets (*HumanEva*). The experimental results show that our method demonstrated higher accuracy and efficiency compared to state-of-the-art approaches.

KEYWORDS

Human Motion Tracking; Multi-Scale; 2D; Monocular Video

1. Introduction

Monocular camera is the most widely and easily available source that records all kinds of human activities. In recent years, many tracking algorithms have been proposed for human motion tracking in 2D monocular videos. A large number of researches are provided in the literature [1] and [2]. Most of 2D human postures tracking frameworks ([3-9]) are inspired by the development of bottom-up pose estimation approaches. Most work tend to focus on building body models or deriving effective detectors in order to tackle such popular problems in tracking as cluttered background or self-occlusions. However, the scale-variation problem is rarely discussed or addressed in most of these approaches. Rather they assume that the person in the video is moving with a rather fixed distance to the camera, resulting in the size of the human figure in the video to be constant, *i.e.*, the perspective scale is fixed. In real life, videos often contain people appearing at any distance to the camera hence appeared in various scales in the videos. Often they are moving towards or away from the camera, resulting in their sizes (scales) to be changed within a video clip. In this paper, we focus on the problem of tracking human motion in multiple scales in monocular image sequences.

A successful approach for 2D human pose tracking in video is to detect the human body and estimate body posture in each frame ('tracking by detection'). For pose estimation in each frame, the pictorial structures model [10,11] is a powerful tool that captures the kinematic relations between parts and allows for exact and efficient inference of the spatial layout of body parts (e.g., [3,8,12,13]). In such models each body part corresponds to a node in a graph, and two nodes are connected by an edge when there is a joint connecting the corresponding body parts. One example is by Ramanan *et al.* [5], who propose to build a color-based specific appearance model based on the detections from a 'stylized pose' detector and then to track the person by detecting the model in each frame. Another system [3] proposes to combine a generic shape-based appearance model with a specific color-based one for human motion tracking. It first estimates rough poses using a shape-based generic part detector and then cluster these estimations based on color information in order to build a specific appearance model.

From experimental results reported in [3], this strategy for learning specific appearance model can achieve better performance than [5]. Although the performance of [3] and [5] is acceptable, a critical problem is that they do not implement the scale-variation issue. Reference [3] states their tracking approach can only track a target at a single scale in a video clip. Reference [5] mentions their system should work theoretically for the multiple scales by searching the pictorial structures over an image pyramid for each frame in the video. But no implementation detail is given in the paper and since it is basically an exhaustive search it should be computationally inefficient.

Recently, there are a few approaches on pose estimation trying to address the scale-variation issue for still images. In [14], an upper-body detector and a foreground highlighting step are used to determine the approximate location and scale information of the person to be tracked. Although it is capable of estimating upper body pose in highly challenging images, the person to be tracked is required to be upright and seen from the front or the back (not the side). Another approach on pose estimation [15] based on generic pose estimation for still images mentioned the scale variation problem. In their approach, the value for the scale parameter is changed within a fixed range at a fixed interval to estimate the human pose in the still images in a trial-and-error fashion. Even though this algorithm is quite generic, it is obviously not attractive for tracking in image sequences because it is cumbersome and computationally inefficient.

In contrast, a more effective approach is proposed in this paper for automatically evaluating and adjusting the scales for the tracked target during the tracking process, which enables the tracking for free-moving human motion with high efficiency and accuracy.

2. Tracking with Adaptively Changed Scale Values

2.1. System Overview

The illustration of our framework is shown in **Figure 1**.

To track the human motion in a video sequence, it is required to detect the person's pose in each frame with a proper scale. A scale checking and adjusting step is incorporated into the tracking process. Two metrics are proposed for detecting and adjusting the scale change. One metric (M_I) is from the height value of the tracked target, which is suitable for some sequences where the tracked target is upright and with no limbs stretching. Generally, since the type of the motion performed by the tracking target is not known, a metric need to be derived to represent the scale changes that is invariant to motion types. For this end, we propose an alternative metric (M_II). Concretely, the images are firstly processed for foreground segmentation which aims to obtain an approximate size of the body blob. This blob size is not used to determine the scale directly. Rather it is used to be compared with the size of the estimated human body (normally in the shape of bounding boxes) from pose estimation to determine whether the scale used for the pose estimation is appropriate. If the comparison shows that the scale value used satisfy the preset condition, the algorithm will proceed to the next frame using the same scale value. Otherwise, the scale value will be adjusted and the frame will be re-processed until the preset condition is met. The metric and condition used for evaluating and adjusting the scale values and how they are incorporated into the tracking framework are detailed in the next subsections.

2.2. Metrics for Scale Estimation

When the tracked person is moving towards or away from the camera, the person appears to be bigger or smaller in the images. Many features could appear differently in the images. In this paper, we propose two types of the metrics used for scale estimation.

2.2.1. Metric from the Height of the Tracked Target: Height_Metric (M_I)

For certain human motions, such as walking, the change in the body height is a good representation for scale variation of the person. A straightforward metric as given in Equation (1) is hence proposed for estimating the scale value in such motions where the tracked target keeps upright with no extreme limbs stretching.

$$s_i = h_i / \alpha \tag{1}$$

where s_i is the scale value used for pose estimation for the i_{th} frame, h_i is the body height in pixels measured from the tracked target in the frame, and α is a reference coefficient, which corresponds to the height of the tracked person in pixels when the scale equals to 1. It is important to note that the scale here is defined to be relative to the value in the training set.

Generally, the scale of the tracked target would not change much between two successive frames in most videos. Based on this fact, we add a necessary rule (Equation (2)) when updating the scale during the tracking process. A small positive constant σ is used as the threshold to check whether the scale to be updated for tracking is suitable or not. In our experiment, we set $\sigma = 0.1$. It means that if the difference between consecutive scales is within 10%, the scale adjustment is acceptable. Otherwise, the scale change is considered to be too drastic to be acceptable, thus avoid error accumulating. It is found through experiments that 10% is a reasonable threshold value when the frame rate is 30 frames per second.

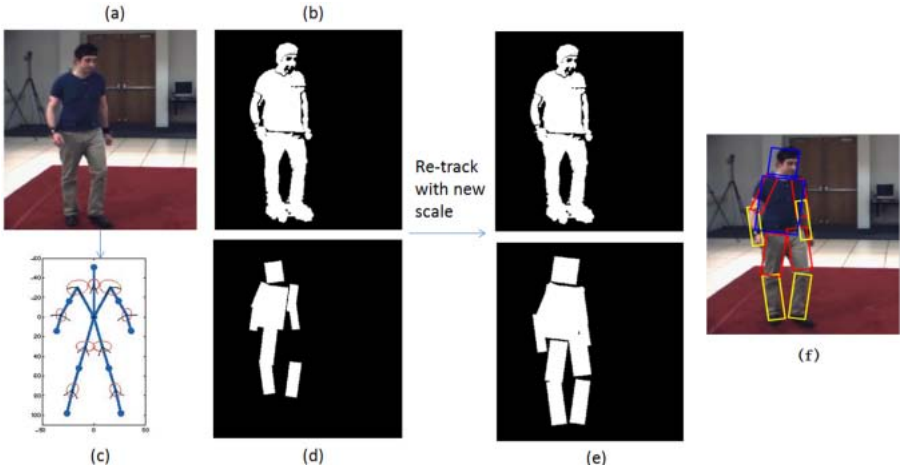

Figure 1. The whole tracking process. (a) the original image (b) the foreground segmentation (c) the kinematic tree model (d) the area of the tracking body parts (represented using bounding boxes) for this frame using one scale value. The pixel numbers from the foreground and from the tracked body parts are counted for scale evaluation. If the scale is deemed inappropriate, the scale will be changed and the frame reprocessed. When the scale is satisfactory, the tracking results is accepted and shown as (f).

$$s_i = \begin{cases} s_i & \text{if } \left|\frac{s_i}{s_{i-1}} - 1\right| \le \sigma \\ s_{i-1} & \text{else} \end{cases} \quad (2)$$

2.2.2. Metric from Pixel Counts: PixelCount_Metric (M_II)

Although the body height is a straightforward metric for sequences such as walking, it is not suitable for motions where the human is not upright. In contrast, the number of pixels the person projected onto an image always changes with respect to the distance between the person and the camera, regardless of the pose/motion of the person. Therefore, the number of pixels occupied by the bounding boxes representing the estimated human is a good indication of the scale value used for pose estimation. The pixel count provides a very good means for estimating and adjusting the scale value. As shown in **Figure 2**, if the two pixel counts are similar (as in row1), it can be determined that the scale used for the pose estimation is acceptable. Otherwise as shown in the second row, the scale value used for pose estimation is far from adequate and needs to be adjusted according to difference between the two pixel counts.

In many situations, the tracked person could stretch his limbs or bend towards/away from the camera, resulting in some parts of the body with more scale changes than the others. Like most state-of-art motion tracking techniques, we do not distinguish the scale differences within body parts since their effect is rather insignificant under the current bounding box framework.

The pixels occupied by the tracked target in images can be obtained through image segmentation, while the pixels occupied by the bounding boxes can be easily

Figure 2. Comparison between the pixel areas of the foreground blobs and the estimated bounding boxes. (a) original image; (b) foreground blobs from image segmentation; (c) bounding boxes from pose estimation; (d) pixel areas covered by the bounding boxes.

identified after pose estimation. Assuming the pixel numbers counted from both operations are denoted as n_1 and n_2 respectively, the scale value used for pose estimation is evaluated and adjusted according to Equation (3).

$$s_i = \begin{cases} s_i & \text{if } \left|r_i - 1\right| \le \sigma \\ \sqrt{r_i} * s_i & \text{else} \end{cases} \quad (3)$$

where s_i is the scale value used for pose estimation for the i_{th} frame, r_i stands for the ratio between n_1 and n_2, and σ is the threshold, again set $\sigma = 0.1$. Since any changes in the scale value will apply to both the width and the height in a 2D image, square root of the ratio r_i is chosen as the coefficient in Equation (3).

An initial value needs to be given for s_i for the first frame. It does not have to be a proper scale, since the pixel numbers n_1 and n_2 counted after image segmentation and pose estimation will be compared to check whether s_1 is an appropriate value. It can be adjusted ac-

cording to Equation (3) and then used for the pose estimation until $r_1 \approx 1$. The updated scale will be used for tracking the second frame, and the same procedure will apply to all remaining frames.

Our framework focuses on the situation using one fixed camera, so the background subtraction is a proper method for image segmentation. In our implementation an extend version of background subtraction [16,17] was selected to provide the blobs of the foreground. Although in general, image segmentation is unable to provide accurate image blobs for representing the tracked target, it is sufficient to provide the approximate size of the human body, and it is easily implemented in our approach.

The pixel number n_2 can be easily obtained by considering the vertices of the resulting bounding box for each body part. The pixels bounded by them can be easily counted with overlapping areas counted only once.

2.3. Tracking

The previous subsections have described how we evaluate and adjust the scale value used for tracking, but do not touch upon how to track a human's body poses. We follow standard approaches of tracking human pose based on human body detection in each frame as described in [11]. Specifically, the human pose tracking is implemented by detecting a strong body model based on pictorial structures [10] in each frame. The pose detector used in this paper is built on [18], which is a generic method for human body part detection and pose estimation.

In our framework, the human body is modeled with 10 decomposed parts: head, torso, left and right lower and upper arms, as well as left and right upper and lower legs. Their configuration is represented as $L = l_1,...,l_{10}$, where the state of part i is given by $l_i = (x_i, y_i, \theta_i, s_i)$ centered at (x_i, y_i) in image coordinates, θ_i is the absolute part orientation, and s_i is the part scale, which equals to the scale value of current frame. Given the image evidence I, the posterior of the part configuration L can be written as

$$p(L \mid I) \propto p(I \mid L)\, p(L). \tag{4}$$

where $p(I|L)$ is the likelihood of the image evidence for the given configuration L; $p(L)$ represents a kinematic tree prior in pictorial structures approach.

The first essential component in the pictorial structures model is the prior $p(L)$, which encodes the kinematic dependencies between the connected parts in the probabilistic form. Such kinematic dependencies in the human body can be captured probabilistically using a tree-structured graphical model. The distribution over configuretion can be factorized into the product as

$$p(L) = p(l_0) \prod_{(i,j)\in G} p(l_i \mid l_j). \tag{5}$$

where G is the set of all edges of connected body parts, l_0 represents the root node in the tree (torso). The probabilities between the dependent child body part l_i and its immediate parent part l_j is denoted as the pairwise terms $p(l_i \mid l_j)$, which are modeled by Gaussians in the transformed space of part joints. The prior for the root node configuration $p(l_0)$ is assumed to be uniform, which allows for a wide range of possible configurations.

The likelihood term $p(I|L)$ in the pictorial structure is the other important component. For simplicity, I_i, which is the image evidence of part i, is assumed conditionally independent given the body configuration L, and each I_i for part i only depends on its own configuration l_i. So the likelihood $p(I|L)$ is decomposed into the product of single part likelihoods

$$p(I \mid L) = \prod_{i=0}^{N} p(I_i \mid L) = \prod_{i=0}^{N} p(I_i \mid l_i). \tag{6}$$

In the implementation, the part likelihoods are modeled with AdaBoost classifiers [19] and the image evidence is represented by a grid of shape context descriptors [20].

The major steps of the proposed multi-scale tracking system are shown in **Figure 1**.

3. Experimental Results

3.1. Evaluation of the Proposed Algorithm

We begin the experimental part of this paper with the evaluation of each proposed metric. Experiments rely on the pre-trained results from *People* dateset of Ramanan [5], which includes persons across a variety of activities.

We quantify the tracking performance of the proposed algorithm by computing the average of correctly localized body parts. One criterion is formally defined in [3]. Let $p(l)$ be the number of pixels located within one rectangle l. Assuming $\hat{l}_m$ is the estimated bounding box and l_m is the ground truth for part m. $\hat{l}_m$ is considered correctly localized if the condition (Equation (7)) is satisfied.

$$\left| p(l_m) \cap p(\hat{l}_m) \right| \geq 0.5 \left| p(l_m) \right| \tag{7}$$

This measure is similar as the PCP ('percentage of correct parts') metric, which originally introduced in [21] and was subsequently used for performance evaluation on pose estimation. So here we named the criterion we adopted as PCP_T.

1) Tracking using the *Height_Metric* (M_I). The proposed multi-scale tracking framework is first applied on two walking sequences from the well-known HumanEva dataset [22], hereby named as *HE_S1_walking* and *HE_S2_walking*. In both sequences, the tracked person walks in a circle, thereby generates image frames with different scales and shows different body orientations including frontal, back, and sideways. **Figure 3** shows

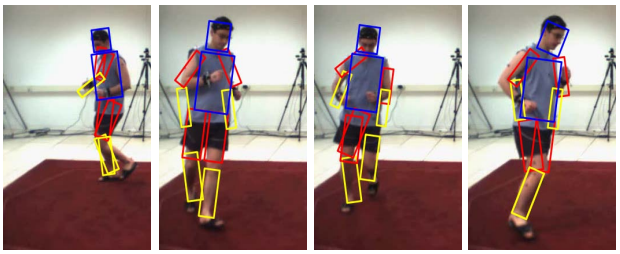

Figure 3. Sample results using the *Height_Metric* (M_I).

sample results on walking sequence from the proposed tracking system.

2) Tracking using the *PixelCount_Metric* (M_II). The proposed tracking framework with *PixelCount_ Metric* is applied on two sequences from the HumanEva dataset [22], hereby named as *HE_Jogging*, and *HE_ Gestures*. The tracked person in videos performs different motions in a circle and shows different scales and body orientations. In addition, the *Gestures* sequence contains several non-lateral motions, such as jumping, kicking, leaning and stretching. The third row of **Figure 4** shows sample results on sequences from the proposed tracking system.

It can be seen from the results that the proposed algorithm can produce satisfactory tracking with varied scales. The tracking results using two different metrics are evaluated with PCP_T criterion and quantitative results are given in **Table 1**.

3.2. Comparison to State-of-The-Art Approaches

The strategy for estimating body's scale described in reference [14] is designed for pose estimation in still images. There is a crucial limitation when directly using it in motion tracking framework, *i.e.*, the person to be tracked is required to be upright and seen from the front or the back (not the side). We implement it *on HE_ Throw_Catch* sequence, in which the tracked target is upright and seen from the front, and compare with our method based on PixelCount_*Metric*. **Table 2** shows the comparison results, which clearly show that the proposed method outperforms the one in [14]. This is due to the fact that the full body detector used in this paper is more reliable than the one in [14].

To further illustrate the performance of the proposed algorithm, we compare it with [3,5] on sequences where the tracked person with and without scale variation, respectively. The reason for choosing these two systems as contrast tests lies in that the basic tracking ideas in [3,5] are similar as ours, such as pictorial structures model, baseline of tracking by detection, *etc.* We implement [3] and [5] based on their provided source code.

We can recall the performance of our proposed framework on walking sequences in section 3.1. The quantitative comparison is listed in **Table 3**. The comparison shows the performance of our framework and [3] is higher than Ramanan's method [5]. This is due to the first two approaches use the shape feature to build the generic model while Ramanan's model bases on the colour feature. Also, *Lu's* method [3] outperforms our proposed system. The reason is that *Lu's* tracking system combines a specific appearance model with the generic detector. That is to say, [3] uses both the shape and colour features, but we only employ the shape feature.

We compare three methods on *HE_Jogging* and *HE_ Gestures*, in which the tracked person appears at different scales. Sample snapshots of the tracking results from these three frameworks are shown in **Figure 4**. The first row is the results by Ramanan's method [5] and the second row shows the results by *Lu's* approach [3]. The results from our approach are shown in row 3. It is obvious that our method can produce satisfactory tracking performance for sequences not only with a wide range of motion types but also with the significant scale varia-

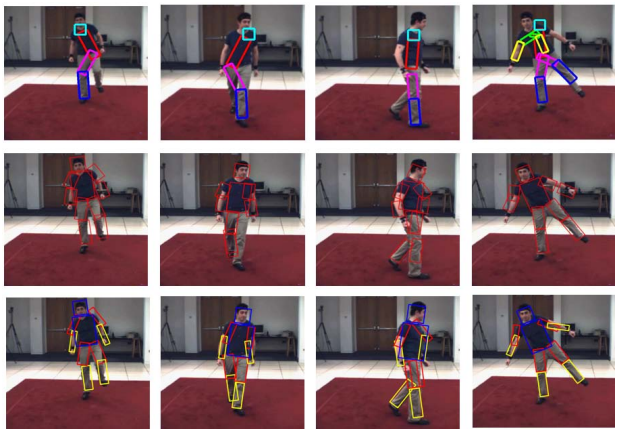

Figure 4. Snapshots of tracking results on sequences with scale variations.

Table 1. Accuracy of tracking results in percentage.

Metric	Torso	Head	Upper leg	Lower leg	Upper arm	Fore arm	Total
M_I	99.2	95.5	85.1	86.5	81.6	79.3	86.4	84.5	82.3	84.1	**86.5**
M_II	98.4	94.5	84.7	83.2	81.8	78.0	81.0	82.5	73.1	72.8	**83.0**

Table 2. Comparison of tracking results on Throw_catch sequence in percentage.

Approach	Torso	Head	Upper leg	Lower leg	Upper arm	Forearm	Total
[14]	95.6	94.8	82.9	78.1	76.9	78.5	70.5	66.5	53.4	52.2	74.9
Proposed	98.5	95.5	85.4	87.8	85.5	83.0	83.9	79.5	73.7	76.3	84.9

Table 3. Comparison of tracking results on Walking sequence with no scale variation in percentage.

Approach	Torso	Head	Upper leg	Lower leg	Upper arm	Fore arm	Total
[5]	94.9	72.3	89.8	51.0	72.0	61.2	13.5	37.5	63.5	48.3	60.4
[3]	100	95.5	87.3	97.5	82.8	91.1	96.9	75.7	86.5	96.5	91.0
Proposed	99.2	95.5	85.1	86.5	81.6	79.3	86.4	84.5	82.3	84.1	86.5

Table 4. Comparison of tracking results on sequences with scale variations in percentage.

Approach	Torso	Head	Upper leg	Lower leg	Upper arm	Fore arm	Total
[5]	52.5	36.3	53.8	65.0	68.8	63.4	24.3	26.7	33.8	42.5	46.7
[3]	91.1	80.3	62.7	65.4	69.8	73.9	60.4	62.6	54.5	52.1	67.3
Proposed	98.4	94.5	84.7	83.2	81.8	78.0	81.0	82.5	73.1	72.8	83.0

tions. The other two approaches fail for frames in which the assumed fixed scale is not suitable. The quantitative comparison is given in **Table 4**, where the accuracy is evaluated based on the tracking results for all frames in the two sequences: *HE_Jogging* and *HE_Gestures*. Clearly, the tracking performance of our approach surpasses [5] and [3], although they perform well for sequences with no scale variation. The tracking for all body parts are remarkably improved. It appears that the method proposed by [5] performs quite poorly when there are significant scale variations in the image sequence. This clearly demonstrates the importance of including scale adjustment in the tracking process, since the overall performance has been greatly improved.

4. Conclusion

In this paper we propose a human motion tracking framework for 2D monocular images, specially address the scale variation problem. An automatic scale evaluating and adjusting algorithm is proposed to adaptively change the scale values during the tracking process. Two metrics for this algorithm are proposed. One is *Height_Metric*, which is a simple and straightforward metric suitable for motions where the tracked target remains upright. The other is *PixelCount_Metric*, which is implemented by computing the ratio between pixel counts of the foreground blobs and the detected body part bounding boxes. This metric is more complicated yet more generic and invariant to motion types. The efficacious of the proposed algorithm is demonstrated through experiments on the publicly available Human Eva datasets, where the proposed algorithm can produce highly satisfactory tracking results.

REFERENCES

[1] R. Poppe, "Vision-Based Human Motion Analysis: An Overview," *Computer Vision and Image Understanding*, Vol. 108, No. 1C2, 2007, pp. 4-18.

[2] H. Zhou and H. Hu, "Human Motion Tracking for Rehabilitations—A Survey," *Biomedical Signal Processing and Control*, Vol. 3, No. 1, 2008, pp. 1-18. http://dx.doi.org/10.1016/j.bspc.2007.09.001

[3] Y. Lu, L. Li and P. Peursum, "Human Pose Tracking Based on Both Generic and Specific Appearance Models," Control Automation Robotics & Vision, 2012, pp. 1071-1076.

[4] J. M. del Rincon, D. Makris, C. O. Urunuela and J.-C. Nebel, "Tracking Human Position and Lower Body Parts Using Kalman and Particle Filters Constrained by Human Biomechanics," *IEEE Transactions on Systems, Man, and Cybernetics Part B*: *Cybernetics*, Vol. 41, No. 1, 2011, pp. 26-37. http://dx.doi.org/10.1109/TSMCB.2010.2044041

[5] D. Ramanan, D. A. Forsyth and A. Zisserman, "Tracking People by Learning Their Appearance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 29, No. 1, 2007, pp. 65-81. http://dx.doi.org/10.1109/TPAMI.2007.250600

[6] X. Lan and D. P. Huttenlocher, "Beyond Trees: Common-Factor Models for 2d Human Pose Recovery," *IEEE ICCV*, Vol. 1, 2005, pp. 470-477.

[7] C. Chang, R. Ansari and A. Khokhar, "Cyclic Articulated Human Motion Tracking by Sequential Ancestral Simulation," *IEEE CVPR*, Vol. 2, 2004, pp. II-45.

[8] D. Ramanan and D. A. Forsyth, "Finding and Tracking People from the Bottom Up," *IEEE CVPR*, Vol. 2, 2003, pp. II-467.

[9] R. Fablet and M. J. Black, "Automatic Detection and Tracking of Human Motion with a View-Based Representation," ECCV, Springer, 2002, pp. 476-491.

[10] M. A. Fischler and R. A. Elschlager, "The Representation and Matching of Pictorial Structures," *IEEE Transactions on Computers*, Vol. 100, No. 1, 1973, pp. 67-92. http://dx.doi.org/10.1109/T-C.1973.223602

[11] P. F. Felzenszwalb and D. P. Huttenlocher, "Pictorial Structures for Object Recognition," *IJCV*, Vol. 61, No. 1, 2005, pp. 55-79. http://dx.doi.org/10.1023/B:VISI.0000042934.15159.49

[12] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient Matching of Pictorial Structures," *IEEE CVPR*, Vol. 2, 2000, pp. 66-73.

[13] D. Ramanan, "Learning to Parse Images of Articulated Bodies," *NIPS*, Vol. 19, 2007, p. 1129.

[14] M. Eichner, M. Marin-Jimenez, A. Zisserman and V. Ferrari, "2d Articulated Human Pose Estimation and Retrieval in (Almost) Unconstrained Still Images," *IJCV*, Vol. 99, No. 2, 2012, pp. 190-214. http://dx.doi.org/10.1007/s11263-012-0524-9

[15] M. Andriluka, S. Roth and B. Schiele, "Discriminative Appearance Models for Pictorial Structures," *IJCV*, 2012, pp. 1-22.

[16] C. Stauffer and W. E. L. Grimson, "Adaptive Background Mixture Models for Real-Time Tracking," IEEE CVPR, Vol. 2, 1999.

[17] P. KaewTraKulPong and R. Bowden, "An Improved Adaptive Background Mixture Model for Real-Time Tracking with Shadow Detection," Video-Based Surveillance Systems, Springer, 2002, pp. 135-144.

[18] M. Andriluka, S. Roth and B. Schiele, "Pictorial Structures Revisited: People Detection and Articulated Pose Estimation," *IEEE CVPR*, 2009, pp. 1014-1021.

[19] Y. Freund and R. E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," *Journal of Computer and System Sciences*, Vol. 55, No. 1, 1997, pp. 119-139.

http://dx.doi.org/10.1006/jcss.1997.1504

[20] S. Belongie, J. Malik and J. Puzicha, "Shape Matching and Object Recognition Using Shape Contexts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 4, 2002, pp. 509-522. http://dx.doi.org/10.1109/34.993558

[21] V. Ferrari, M. Marin-Jimenez and A. Zisserman, "Progressive Search Space Reduction for Human Pose Estimation," *IEEE Conference on CVPR*, 2008. pp. 1-8.

[22] L. Sigal, A. O. Balan and M. J. Black, "Humaneva: Synchronized Video and Motion Capture Dataset and Baseline Algorithm for Evaluation of Articulated Human Motion," *IJCV*, Vol. 87, No. 1, 2010, pp. 4-27. http://dx.doi.org/10.1007/s11263-009-0273-6