# Human Body Tracking and Joint Angle Estimation from Mobile-phone Video for Clinical Analysis

Jehoon Lee, Peter Karasev, Liangjia Zhu, and Allen Tannenbaum
School of Electrical and Computer Engineering
Georgia Institute of Technology, Atlanta, Georgia, USA
{jehoon.lee, pkarasev, ljzhu}@gatech.edu, tannenba@ece.gatech.edu

## Abstract

*This note introduces a model-free and marker-less approach for human body tracking based on a dynamic color model and geometric information of a human body from a monocular video sequence. A multivariate Gaussian distribution is learned online from sequential frames to represent the non-stationary color distribution. Images are first filtered according to the current color model allowing a human body to be segmented from a background. Next, the segmented image is partitioned based on geometric prior knowledge of human structure and each body part of interest is separately tracked. Finally, eigenaxis based joint angle estimation is carried out to evaluate jumping and landing posture. The resulting data facilitates motion analysis of a specific set of clinically interesting quantities that allow post-operative evaluation and correlation of injury statistics with a subject's mechanics. The proposed approach is tested on the different views of human jumping and landing sequences in a noisy and cluttered environment with video from a mobile-phone camera. Experimental results demonstrate the practical applicability and robustness of the proposed algorithm in tracking human motion captured with a monocular camera system.*

## 1 Introduction

Tracking a human body's joint angle enables important clinical analysis for post-operative analysis and prediction of a healthy subject's likelihood of injury [1]. To enable such video-based estimation in a minimally controlled environment with consumer-grade hardware, we present a model-free and marker-less joint angle tracking and estimation method. By restricting the tracking task to the body segments of interest and employing a dynamic online estimate of the color distribution, fast and robust tracking results are obtained from a low signal-to-noise-ratio monocular camera source.

A vast number of tracking methods for a human body are proposed in literature [2]. Most commonly, they adopt a explicit body model, such as a model built on a kinematic chain consisting of conical sections [3] or a model composed of a set of kinematic and tapered cylinder shape [4]. To optimize the map of image-features to model, the work in [3] introduced annealed particle filters suitable for high dimensional configuration spaces. In [4], data-driven MCMC technique is proposed to estimate 3D poses showing the best model match. In contrast, our method is carried out without explicitly using a body model and markers that can indicate and describe a human figure. The motivating use-case for our proposed algorithm is a hand-held camera of a mobile phone in a home or office setting. Accordingly, only a monocular camera can be used; approaches using multiple cameras [5] are outside the scope of feasibility in common daily environments.

Noise and clutter in such an uncontrolled setting further complicates the already non-trivial task of human body segmentation. Region-based active contours are a prevalent method for image segmentation in the presence of noise [6]; they are generally formulated to evolve by gradient descent. Methods based on active contour models show a robust result in segmenting a region of interest from noisy backgrounds. While this approach will in many cases robustly segment a desired region of interest from a cluttered background, it is highly sensitive to initialization and can easily become stuck in a local minimum. In addition, this framework is not suitable in tracking a human body with fast and blurred motion and dealing with the disappearance of some parts of a human body due to narrow field-of-view of a phone camera. Therefore, we employ active contour models for initialization of our tracking framework. Then, a dynamic color-model based segmentation is proposed to properly segment a human body and maintain the track.

Segmenting objects accurately and reliably requires a robust color model. Modeling the skin-color is a challenge that must be addressed in marker-less human body segmentation: numerous background objects, such as a wooden desk or white wall, in tandem with indoor lighting, make skin regions very similar in pixel-value to portions of the background. During a video sequence, the subject's motion can dramatically change the apparent pixel-values due to overhead lights being obscured by their arms, head, and torso; even for a specific human and environment, the observed color values can fluctuate significantly. Most methods for the construction of a skin-color model that robustly characterizes human skin under varying lighting conditions and skin tones rely on supervised training with several images before segmentation begins [7]. Naturally such an approach has limitations: collecting and training with example images of skin is a time-consuming task and cannot completely cover the possible range of a skin-color distribution. We propose an online adaptive pixel-color model which captures the time-varying distributions of skin and clothing color.

## 2 Proposed Algorithm

### 2.1 Human Body Segmentation and Partitioning

The RGB (Red, Green, and Blue) representation of color images is most common but is not suitable for characterizing a skin-color region because the chrominance and luminance information are correlated in the

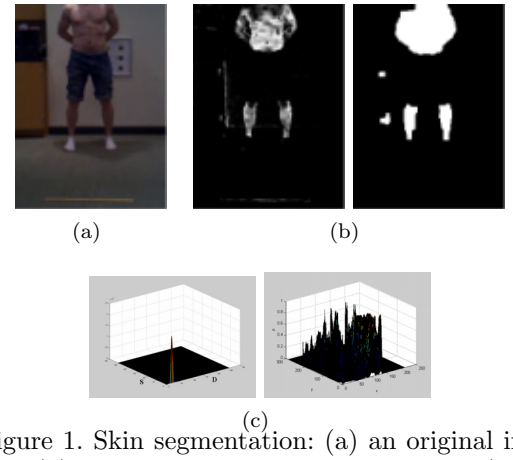


Figure 1. Skin segmentation: (a) an original image. (b) a result using the proposed method (left) and its smoothed result by a morphological filter (right). (c) a skin color model (left) and the likelihood of skin for the image (right).

RGB space. Luminance has an effect on skin color according to the varying levels of brightness. To overcome this, the robustness of several color spaces is analyzed in [8]. For example, normalized RGB showing chromatic colors is effectively used for skin color segmentation [9]. In this present work, the combination space of a difference space (D) between a red space (R) and a green space (G), and a saturation space (S) of a color image is used. This space provides a good indicator for some colors. In particular, skin color is well classified and its distribution resides within $[15, 90]$ in a D space via our empirical observations. We denote by $I$ the image, by $I_{(\cdot)}$ the associated component (or space) of an image $I$. The image used in this work is defined as:

$$I_D = I_R - I_G, \tag{1}$$
$$I_S = \frac{\max(I_R, I_G, I_B) - \min(I_R, I_G, I_B)}{\max(I_R, I_G, I_B)}.$$

The proposed color model is statistically designed as a multivariate Gaussian distribution. The likelihood of color for a pixel, $\mathrm{x} \in \mathbf{R}^2$, with a pair value of $\mathrm{m} = [d, s]^T$, $d \in I_D$, $s \in I_S$, is obtained from the color model, which is given by:

$$p(\mathrm{x}|\mathrm{m}) = |\Sigma_c|^{-\frac{1}{2}} \exp -\frac{1}{2}(\mathrm{m} - \mu)^T {\Sigma_c}^{-1}(\mathrm{m} - \mu), \tag{2}$$

where $\mu \in \mathbf{R}^2$ is a mean value vector for D and S spaces and $\Sigma_c \in \mathbf{R}^{2\times 2}$ is a covariance matrix. The pixels from equation (2) are labeled as a foreground image of interest if they satisfy $p(\mathrm{x}|\mathrm{m}) > c_{th}$. Here $c_{th}$ is a user-defined constant for the threshold and is empirically selected as $[0.3, 0.5]$ in our experiments. Therefore, it produces an binary image, $b(\mathrm{x})$, representing the segmented region as 1 and backgrounds as 0. The binary image is smoothed by a morphological filter, $f_s(\cdot)$: $\bar{b}(\mathrm{x}) = b(\mathrm{x}) \otimes f_s(\cdot)$. The size of the filter depends on the noise present in the sequence. This allows improved segmentation results in a highly cluttered environment. Since $\bar{b}(\mathrm{x})$ can be composed of several pixel groups, for convenient notation, we let $g^i(\mathrm{x})$ be each labeled pixel group and its region is denoted by $R^i$. Figure 1 shows segmentation results of skin color by using the proposed method.

After segmentation based on the color model, the labeled regions are selected as each body part of interest among several labeled regions by considering the centroid and the size of the previously segmented body parts as follows:

$$E_c{}^\star = arg \min_{g^i(\mathrm{x})} E_c^i, \tag{3}$$

where

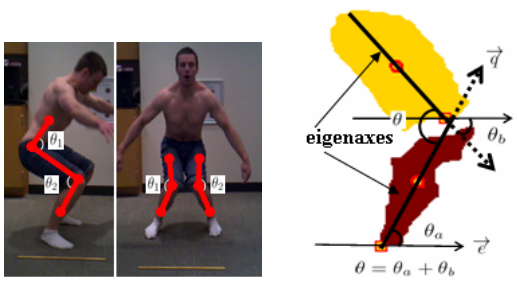
$$E_c^i = \|(c_\mathrm{x}^i - c_\mathrm{x}(0))^2 + \left(\int_{\mathrm{x}\in g^i} \mathrm{x}d\mathrm{x} - \int_{\mathrm{x}\in g(0)} \mathrm{x}d\mathrm{x}\right)^2 \|. \tag{4}$$

Here $\|\cdot\|$ is the Euclidean norm and $c_\mathrm{x}^i \in \mathbf{R}^2$ is the centroid of each labeled region $R^i$ at a current frame. $c_\mathrm{x}(0)$ and $g(0)$ is a centroid and a labeled group of the previously selected region, respectively. The centroid of a labeled region is computed by the average value of the points' coordinates inside the region $R^i$ [10]. A labeled region with the minimum energy $E_c$ is selected as a region of interest, such as torso, upper and lower legs.

## 2.2 Joint Analysis

The segmented human body can provide a human motion analyst with crucial information, such as the joint angle between two pieces of limbs. The proposed method is designed to help an orthopedist study healthy subjects and see the probability of injury based on posture of jumping and landing and to study recovery in post-operative patients. To this end, joint angles of a human jumping and landing are estimated in a daily environment where professional medical facilities are not available. Some joint angles of interest in this work are shown in Figure 2 (a); angles between knees and toes, and angles between a knee and a hip. Some minimal control of the experimental setup is assumed. First, the human is instructed to jump either along a line in the camera plane *side-view* or facing the lens *front-view*. With the narrow angle-of-view in the mobile-phone camera source, lens distortion is negligible. The individual is wearing some specific clothing (*e.g.* homogenously colored pants). There is no constraint on a specific color or reflective markers.

In this work, eigenaxes are used for estimating joint angles between body parts because they provide orientation information of the shape derived from a binary pixel group $g^i$. Eigenaxes are defined as the eigenvectors of a covariance matrix of all vectors representing the point coordinates inside $R^i$ [10]. Figure 2 (b) illustrates the method of calculating the angle between a upper leg and a lower leg using a major principle axis. In Figure 2 (b), the angle is simply calculated by using trigonometric functions: $\theta_{(\cdot)} = \cos^{-1}(\overrightarrow{q} \cdot \overrightarrow{e}) \times \frac{180}{\pi}$ where $\overrightarrow{q}$ and $\overrightarrow{e}$ are a major priciple axis vector and a horizontal axis vector, respectively. A joint position is obtained from determining the intersection point of boundaries of $g^i$, the line defined with inclination $\theta_{(\cdot)}$, and a centroid $c_\mathrm{x}^i$.




(a) (b)

Figure 2. (a) joint angles of interest at a side view (left) and a front view (right). (b) joint angle estimation using eigenaxes

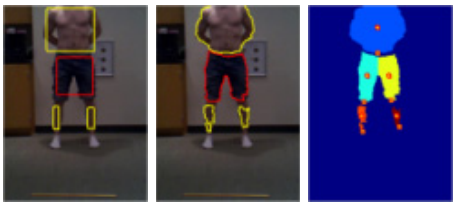


Figure 3. Initialization process of tracking: (From left to right) initial contours, final contours, and body part separation and interesting point detection, i.e., circle points and square points indicate the centroid of each body part and joint points of interest, respectively.

### 2.3 Human Motion Tracking

Initialization to extract the color information of a human body that will be tracked is manually achieved at the first frame of a sequence. To this end, one can use any type of segmentation methods, such as region growing, or active contour models. In our work, we use a region-based active contours driven by Bhattacharyya gradient flow described in [6]. To use this, it is assumed that the human body of interest is fully viewed within a camera view angle at the first frame. To generate the color models for each color, initial contours are carefully located so that they evolve over objective color regions. Figure 3 shows an initialization process at the first frame of a sequence shown in Figure 4. Here two color models are generated. They are re-generated online for each frame during a sequence to capture the time-varying distributions of colors, which is based on the result of the previous frame. The proposed tracking framework is described as follows:

a. Initialize active contours and evolve them to define regions of interest at $t = 0$ where $t$ is a time step.

b. Generate color models for each color region and filter the image from equation (2): $p(\mathrm{x}, t = 0|\mathrm{m})$.

c. Segment and label a pixel group: $g^i(t)$.

d. Compute the energy $E_c^i$ for each labeled region $g^i$ by using equation (4) to decompose the segmented human body into each body part of interest:

$$E_c^i(t) = \|(c_{\mathrm{x}}^i(t) - c_{\mathrm{x}}(t-1))^2 + \left(\int_{\mathrm{x}\in g^i(t)} \mathrm{x}d\mathrm{x} - \int_{\mathrm{x}\in g(t-1)} \mathrm{x}d\mathrm{x}\right)^2 \|. \quad (5)$$

e. Update color models based on the previously selected regions of each body part $g^i(t-1)$: $p(\mathrm{x}, t|\mathrm{m})$

f. Go to step c: $t = t + 1$.

## 3 Experiments

The proposed algorithm was tested on sequences of human jumping and landing showing front and side views of human motion in real environments. All sequences are taken from a low-resolution, small-aperture, narrow field-of-view monocular camera built into a mobile phone in cluttered backgrounds. Due to these properties of the used mobile-phone camera, the acquired images have high noise and low quality. In addition, some parts of the tracked human body (in particular, the torso) are out of camera view in some frames due to a narrow angle lens.

Figure 4 and 5 show the experimental results of the proposed algorithms for front and side views of fast jumping and landing sequences, respectively. Acceptable segmentation and tracking results are obtained and joint angles of interest are estimated to analyze jumping and landing posture. Note that the quality of images is poor and the tested sequences show fast motion. To highlight the effectiveness of the proposed dynamic color model in capturing the time-varying color distributions, results by using a static color model acquired at the first frame are also shown in Figure 4. Figure 6 displays the graphs of joint angles of interest, $\theta_1$ and $\theta_2$, over the sequences shown in Figure 4 and 5, respectively.

## 4 Conclusion

In this note, we have shown an effective algorithm to track a human body and estimate its joint angles by incorporating dynamic color-model based segmentation and eigenaxis based angle estimation. No explicit human model and markers is needed in the proposed framework. The results of experiments showed robust performance of the proposed algorithm and applicability of a joint angle analysis. However, the proposed algorithm has some limitations that we intend to overcome in our future work. Since the proposed algorithm largely depends on the segmented body parts to extract their axis points and estimate joint angles, it is difficult to guarantee the accuracy of joint angles due to missing or inaccurate segmentation. Thus, one can employ a 2D or 3D body model to complement poor segmentation for the improvement of angle estimation.

## References

[1] M. Jagodzinski, V. Kleemann, P. Angele, V. Sch "onhaar, KW Iselborn, G. Mall, and M. Nerlich, "Experimental and clinical assessment of the accuracy of knee extension measurement techniques," *Knee Surgery, Sports Traumatology, Arthroscopy*, vol. 8, no. 6, pp. 329–336, 2000.

[2] T.B. Moeslund and E. Granum, "A survey of computer vision-based human motion capture," *Computer Vision and Image Understanding*, vol. 81, no. 3, pp. 231–268, 2001.

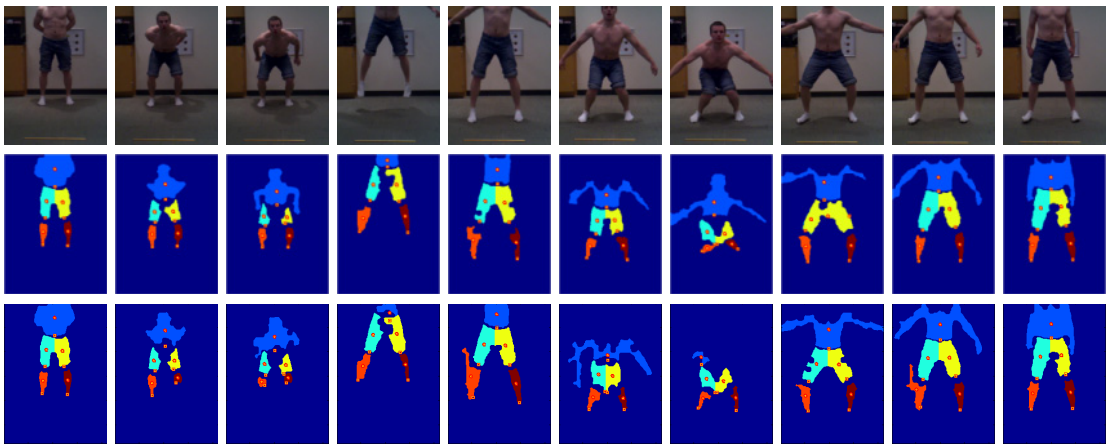


Figure 4. Sequence I: *front-view* jumping and landing posture. The torso partially disappears while jumping as a consequence of narrow angle-of-view. Incident overhead lighting greatly varies during the sequence. Bottom row: The results by using a static color model.

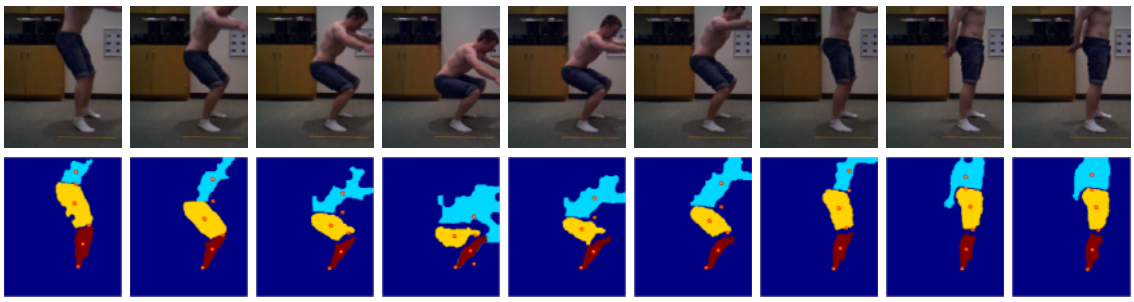


Figure 5. Sequence II: *side-view* jumping and landing posture. There is a disappearance of arms and head from the frame during the sequence. The dynamic color model adapts to observed pixel values.

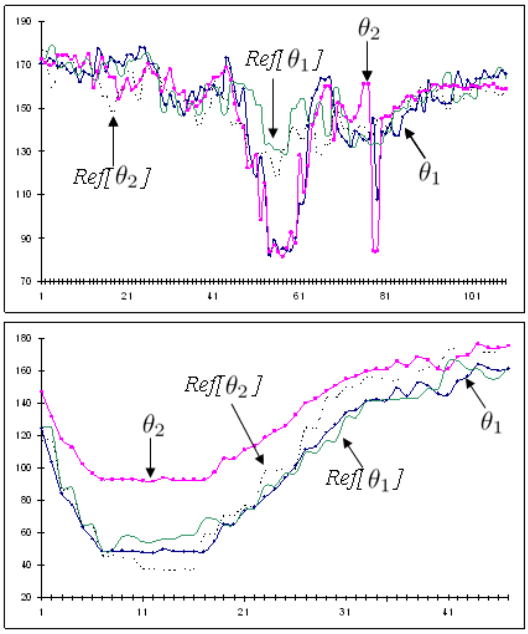



Figure 6. Graphs of estimated angles of the sequences of Figure 4 (top) and the sequences of Figure 5 (bottom). A circled line and a squared line denote the angle of $\theta_1$ and $\theta_2$ of each viewpoint, respectively. Their reference angles are denoted by a solid line and a dotted line, respectively. Extrema of the angles $\theta_1$ and $\theta_2$ encode *maximal knee flexion* [1].


[3] J. Deutscher, A. Blake, and I. Reid, "Articulated body motion capture by annealed particle filtering," in *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 2000, pp. 126–133.

[4] M.W. Lee and I. Cohen, "A model-based approach for estimating human 3D poses in static images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 905–916, 2006.

[5] C. Bregler and J. Malik, "Tracking people with twists and exponential maps," in *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 2000, pp. 8–15.

[6] O. Michailovich, Y. Rathi, and A. Tannenbaum, "Image segmentation using active contours driven by the Bhattacharyya gradient flow," *IEEE Transactions on Image Processing*, vol. 16, no. 11, pp. 2787–2801, 2007.

[7] S.L. Phung, A. Bouzerdoum, and D. Chai, "Skin segmentation using color pixel classification: analysis and comparison," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 148–154, 2005.

[8] P. Sebastian, Y.V. Voon, and R. Comley, "The effect of colour space on tracking robustness," in *Industrial Electronics and Applications, 3rd IEEE Conference on*, 2008, pp. 2512–2516.

[9] M. Wimmer, B. Radig, and M. Beetz, "A person and context specific approach for skin color classification," in *Pattern Recognition, 18th International Conference on*. IEEE, 2006, vol. 2, pp. 39–42.

[10] L. da Fontoura Costa, *Shape analysis and classification: theory and practice*, CRC press, 2001.