인체 추적을 위한 구성요소 기반 확률 전파

Component-based density propagation for human body tracking

신 영 숙* 차 은 미** 이 경 미***
Young-Suk Shin EunMi Cha Kyoung-Mi Lee

요 약

본 논문에서는 구성요소와 그들 간의 유연한 연결을 가진 인체 모델로 인체를 추적하기 위한 구성요소 기반 확률 전파를 제안한다. 인체는 사람의 동작을 추적하는데 필요한 6개의 인체 부위로 나뉘는데, 머리, 몸통, 왼팔, 오른팔, 왼발, 오른발 등이다. 제안하는 추적 방법은 인체 전체의 실루엣을 추적하지 않는 것이 아니라 구성요소로 이루어진 인체모델을 이용하여 인체의 각 부위를 개별적으로 추적하게 된다. 제안된 인체 추적 시스템은 유아의 동작 교육에 적용되는데, 균형잡기, 앙감질, 뛰기, 걷기, 회전하기, 구부리기, 뻗기와 같은 동작을 추적하는데 이용된다. 제안하는 시스템은 인체 모델의 각 부위를 개별적으로 탐지하고 움직임을 추적하여 평균 97%의 추적율을 획득하였다.

Abstract

This paper proposes component-based density propagation for tracking a component-based human body model that comprises components and their flexible links. We divide a human body into six body parts as components - head, body, left arm, right arm, left foot, and right foot - that are most necessary in tracking its movement. Instead of tracking a whole body's silhouette, using component-based density propagation, the proposed method individually tracks each component of various parts of human body through a human body model connecting the components. The proposed human body tracking system has been applied to track movements used for young children's movement education: balancing, hopping, jumping, walking, turning, bending, and stretching. This proposed system demonstrated the validity and effectiveness of movement tracking by independently detecting each component in the human body model and by acquiring an average 97% of high tracking rate.

☞ keywords: 인체 모델, Density 전파, 구성요소, 2D 사람 추적

1. Introduction

Tracking a human body has become an important issue in computer vision. There are various areas of applications for tracking the human body, in the medical field for analyzing person's walk, the virtual reality field for avatar controls, the automation field for interacting between people and machines, the surveillance field for security, and the physical education field for analyzing ballet motions and gymnastics [2,3].

Many researchers have tried different methods for reliable human body tracking. Lowe estimated human body movements using robust dynamic models from video [5]. Mittal *et al.* estimated human body poses using two-dimensional (2-D) silhouette shape analysis [6]. And Rosenhan *et. al* proposed a silhouette-based human motion estimation system achieved more accurate results than using marker systems [7].

* 정 회 원 : 조선대학교 정보통신공학과 교수
ysshin@chosun.ac.kr
** 준 회 원 : 덕성여자대학교 대학원 교육영상매체학과 석사과정 scott5@duksung.ac.kr
*** 종신회원 : 덕성여자대학교 컴퓨터공학부 교수
kmlee@duksung.ac.kr
[2008/01/25 투고 - 2008/01/29 심사 - 2008/04/25 심사완료]

Table 1. Survey on component-based tracking

	proposed method	[9]	[10]	[12]	[13]
what to track	Full body parts	Head, Hands, Feet	Head, Body, arm, leg	Head, Torso, Legs	Upper and lower arm
how to estimate	density propagation	Kalman filter	Anisotropic Gaussian filter	Inferencing stochastic grammars	skin color and inverse kinematics
others	apply avatar	apply avatar	walking	human action	human body action
year	2008	1998	2007	2001	2005

Still, tracking a human body is made particularly difficult by its complex appearance and motion. A full-body approach is subject to a high degree of transformation because the method is affected significantly by changes in posture, lighting conditions, overlapping, transformation, noise, and incompleteness of feature extraction. Even for the same person and the same movement, it is not straightforward to track the whole human body and to recognize precise movement. Because of such partial distortions, the human body should be considered as a set of body parts and tracked independently. A component-based approach considers human body parts as components and subsequently uses the relationships between these components to represent the entire human body. Table 1 shows existing component-based tracking of a human body. Alex *et al.* found information on the location of head, hands and feet using Kalman filter from five cameras and applied the location information to each component of an avatar [9]. Lee and Ko tracked a head, a body, a left arm, a right arm, a left leg and a right leg using anisotropic Gaussian Filter [10]. Cho and Cho proposed the inferencing stochastic grammars for recognizing human actions [12]. Park *et al.* found the location of face and arms using skin color and inverse kinematics [13]. So far, however, component- based human body tracking is applied to only some parts of the human body, such as the upper body, and not the whole human body.

Recently, density propagation has been adopted for human body tracking. Even if the human motion does not track clearly in a current frame, density propagation can estimate the position of human through the previous frames. Wang and Leow articulated human body posture using non-parametric Belief propagation from single or multiple images [8]. Bernier and Cheung-Mon-Chan tracked a human upper-body pose by Particle filtering, and mapped the video frames into a three-dimensional (3-D) model, by Belief propagation [1]. The human body can be expressed by a whole human body's silhouette, and tracked by using density propagation [4].

In this paper, we propose component-based human body tracking using density propagation. A human body is modeled by dividing each part of the entire human body into components. Instead of tracking a whole body, we track, as components, independent body parts.

2. Component-based human body model

2.1 Component-based human body model

In this section, we introduce a component-based human body model that connects body parts. The human body model consists of six body components(i) – head, body, left arm, right arm, left foot, and right foot – that are most important in tracking the human body. Each component contains geometric information (position, relative size, and shape) and appearance information (color). Each component also includes information on its link to other components, such as which of the four sides that represent each component are connectible, the names of the parts being connected, the connecting angles and the connecting distances:

$$Human = (C_i, R_i), \qquad i = 1, \cdots, 6 \tag{1}$$

where $Human$ represents a human body model that consists of I(=6) number of components, while C_i refers to each component geometric and appearance information, and R_i represents information on the link between components.

2.2 Component-based human detection

Each component should be detected and initialized for human body tracking. Given an image frame (Fig.1(a)), we subtract a background image to segment a foreground human body image. We assume that a foreground image (Fig.1(b)) is a semantically binary image consisting of foreground pixels and background (white) pixels.

Such a binary image is grouped according to connected foreground pixels and segmented regions. Then the labeled connected components are filtered out depending on the bounding box of the region, by removing noise-like regions and non-human body part-like regions.

- regions which are too long for a side of the bounding box,
- regions having small areas,
- regions having a small values of elongatedness,

$$\frac{\text{a shorter side of the region}}{\text{a longer side of the bounding box}}$$

- regions having a small value of compactness

$$\frac{\text{area of the region}}{\text{area of the bounding box}}$$

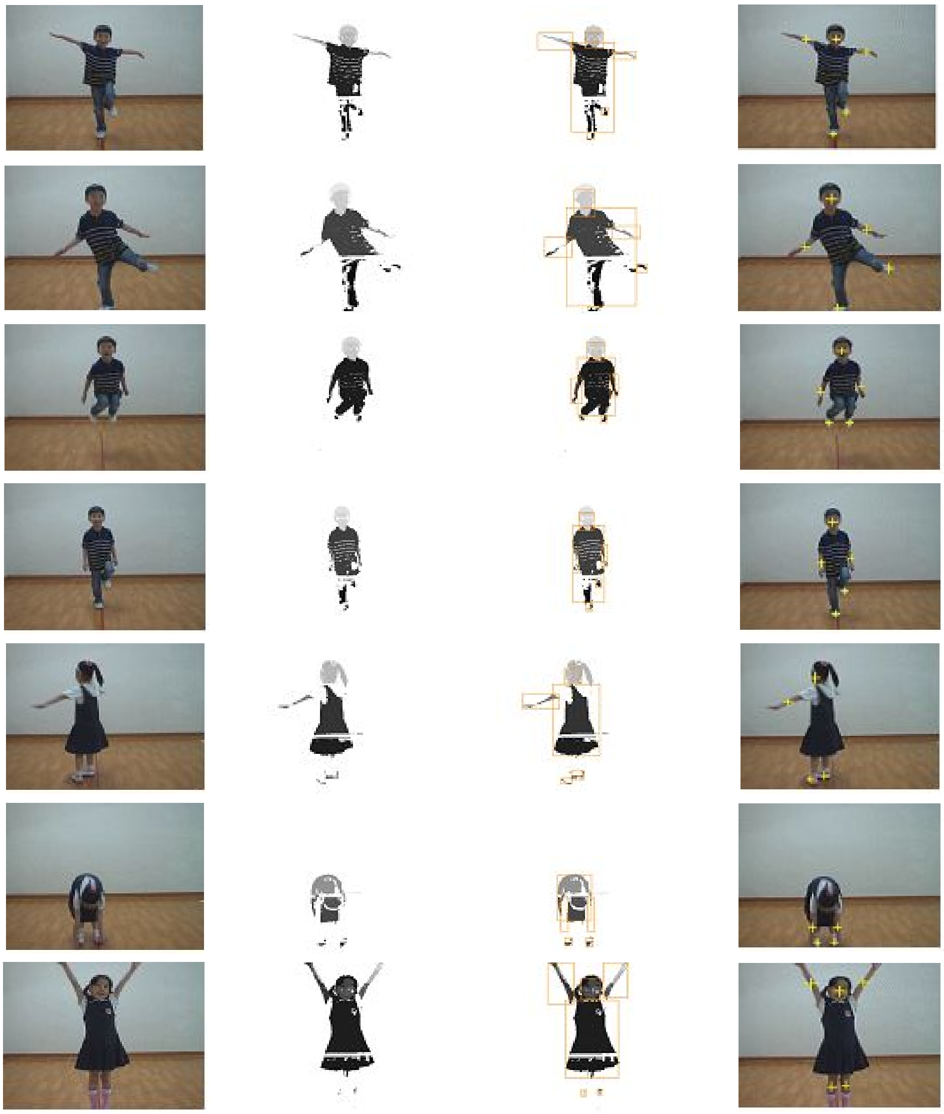
The above criteria provide a set of candidate body component areas without scanning all possible aspects ratios and possible positions into binary segment areas. This improvement leads to greater precision in terms of locating components and helps to detect the body from the background, especially when parts of the surrounding background are in motion.

Then, the grouped foreground is initialized as body components using the human body model in Sec.2.1. The human body model detects the color data of skin or the assumed test clothes if information on clothes is given. In this paper, the face, the left arm and the right arm can be detected using skin color, and the left and the right feet detected by slippers'color because children in the kindergarten that made the video were all wearing slippers. The remaining part is then defined as a torso (Fig.1(c)). We compute the center of each component for probabilistic tracking using the density propagation (Fig.1(d)).

Fig. 1 shows the component-based detection of seven movements – balancing, hopping, jumping, walking, turning, bending and stretching. Fig.1(b) shows a segmented human body. Fig.1(c) shows detected components after removing unnecessary areas and Fig.1(d) shows the centers of components.

3. Component-based density propagation

In this section we track each component in the current frame using the information acquired from the previous frames. To track each component we



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

Fig.1. Detection of components in a human body: (a) input video frame, (b) segmenting foreground, (c) detecting each component, and (d) centering each component. There are seven movements - balancing, hopping, jumping, walking, turning, bending, and stretching, from the top row to the bottom row.

propose to find the location information and relationship among components using the density propagation.

3.1 The Bayes' theorem

The probability theory is the academic approach on coincidence and uncertain. This theory is

positively necessary in pattern recognition dealing with recognition for case the incident occurred in uncertain situations. If there are two events w and x, the conditional probability of the event w, $P(w|x)$, is defined as the probability that the event w will occur given the knowledge that the event x has already occurred. If the events w and x are not independent, then the probability of the intersection of w and x is defined by $P(w \cap x) = P(w|x)P(x) = P(x|w)P(w)$.
From this definition, the conditional probability $P(w|x)$ is easily obtained :

$$P(w|x) = \frac{P(x|w) \cdot P(w)}{P(x)} \qquad (2)$$

The Bayes' theorem is an important statistical approach for pattern recognition problem, especially when mathematically dealing with decision problem in uncertain situations. Suppose that the event class w is divided into a collection of N mutually exclusive events (sets). The event x can be written as the union of the N disjoint events, $x = xw_1 + xw_2 + \cdots + xw_N$. This implies total probability theorem:

$$P(x) = P(x|w_1)P(w_1) + P(x|w_2)P(w_1)P(w_2) + \cdots + P(x|w_N)P(w_N)$$

The total probability theorem and the definition of the conditional probability may be used to derive Bayes theorem:

$$P(w_j|x) = \frac{P(x|w_j) \cdot P(w_j)}{\sum_{k=1}^{N} P(x|w_k) \cdot P(w_j)} = \frac{P(x|w_j) \cdot P(w_j)}{P(x)} \qquad (3)$$

where w_j is a class *j*-th and x is a feature vector. $P(w_j)$ is the prior probability of the class w_j and $P(x)$ is the probability to detect x. $P(w_j|x)$ is the posterior probability for the class w_j, if an observation x is given. $P(x|w_j)$ is the conditional probability to detect the observation x, if the class w_j is given.

3.2 Density propagation

The density propagation is the estimate of an object in frame *k*, based on the observation up to frame *k-1*, as in the Particle filter or the Kalman filter [4]. It is necessary to track and match an object using the density propagation to obtain the probability of tracking the object, $P(X_k | Z_{k-1})$.

If the probability density of an object to be tracked up to frame $k-1$, $P(X_{k-1} | Z_{k-1})$, is provided, the probability density of an object to be tracked in frame *k*, $P(X_k | Z_{k-1})$, can be estimated by propagating the information of the object tracked up to frame $k-1$. The probability density $P(X_k | Z_{k-1})$ can be obtained by applying the sample propagated from $P(X_{k-1} | Z_{k-1})$. The density propagation can be expressed by changing Eq.(3) as follow:

$$P(X_k | Z_{k-1}) = \frac{P(Z_{k-1} | X_k)P(X_k)}{P(Z_{k-1})} \qquad (4)$$

where $P(X_k)$ indicates a prior probability of an object that can be estimated in frame *k*, and $P(X_k | Z_{k-1})$ indicates a posterior probability of an object that can be estimated in frame *k*, when the concerned object tracked up to frame $k-1$ is provided. The probability $P(Z_{k-1} \mid X_k)$ indicates a conditional probability to be decided as the concerned object in frame $k-1$, when an object is tracked in frame k. The probability $P(Z_{k-1})$ indicates a probability of an object in frame $k-1$.

3.3 Component-based density propagation

Component-based density propagation can track each component through the tracked information in human motion from frame $k-1$ to the current frame k using each component geometric and appearance information, C_i. And the connection information of each component, R_i in Sec.2.1. If the probability of a component tracked up to frame $k-1$, $P(X_{k-1}^{(C_i,R_i)} \mid Z_{k-1}^{(C_j,R_j)})$, is given, the probability of a component to be tracked in frame k, $P(X_k^{(C_i,R_i)} \mid Z_{k-1}^{(C_j,R_j)})$, can be estimated by propagating the information of tracked each component. By applying propagated samples from the probability $P(X_{k-1}^{(C_i,R_i)} \mid Z_{k-1}^{(C_j,R_j)})$, the probability of correctly identifying the component $P(X_k^{(C_i,R_i)} \mid Z_{k-1}^{(C_j,R_j)})$ can be obtained by changing Eq.(4) as follows:

$$P(X_k^{(C_i,R_i)} \mid Z_{k-1}^{(C_j,R_j)}) = \frac{P(Z_{k-1}^{(C_j,R_j)} \mid X_k^{(C_i,R_i)})P(X_k^{(C_i,R_i)})}{P(Z_{k-1}^{(C_j,R_j)})} \quad (5)$$

where I and J consist of six components, respectively – head, body, left art, right arm, left and right feet. $P(X_k^{(C_i,R_i)})$ indicates a prior probability of components that can be tracked from frame k and $P(X_k^{(C_i,R_i)} \mid Z_{k-1}^{(C_j,R_j)})$ indicates a posterior probability of the components in frame k, when components tracked up to frame $k-1$ are given. The probability $P(Z_{k-1}^{(C_i,R_i)} \mid X_k^{(C_j,R_j)})$ indicates the conditional probability of the concerned components in frame $k-1$, when components are tracked in frame k. The probability $P(Z_{k-1}^{(C_j,R_j)})$ indicates the probability of components tracked up to frame $k-1$.

Then, it is required to match components tracked by density propagation. To measure the similarity between components, we use the Euclidean distance:

$$\text{distance}(X_k^{(C_i,R_i)}, Z_{k-1}^{(C_j,R_j)}) = \sqrt{(X_k^{C_i} - Z_{k-1}^{R_i})^2 + (X_k^{C_j} - Z_{k-1}^{R_j})^2} \quad (6)$$

where C_i indicates a geometric and appearance information of a component i and R_i is a connection information of component i. $X_k^{(C_i,R_i)}$ is the component i tracked in frame k. $Z_{k-1}^{(C_j,R_j)}$ is the component j tracked up to frame $k-1$. For the distance of C_i and R_i in frame k from C_j and R_j in frame $k-1$, D_{C_i}, which is the smallest value, can be decided as the concerned components. In such a manner, the components repeat the matching process while every frame is coming.

$$D_{C_i} = \min(\text{distance}(X_k^{(C_i,R_i)} - Z_{k-1}^{(C_j,R_j)})) \quad (7)$$

The probability $P(X_k^{(C_i,R_i)} \mid Z_{k-1}^{(C_j,R_j)})$ of the concerned components (Eq.(5)) can be combined as the sum of weights. In this paper, each component can be tracked in each frame:

$$\sum_{i=1}^{I}\sum_{j=1}^{J} w_i P(X_k^{(C_i,R_i)} \mid Z_{k-1}^{(C_j,R_j)}) \quad \text{and} \quad w_i = 1 - \frac{D_{C_i}}{\sum_{m=1}^{M} D_{C_m}} \quad (8)$$

where a weight w_i is the value between 0 and 1, and the sum is 1. D_{c_i} indicates information decided as the concerned component i and D_{c_m} can be indicated as the sum of information decided as the concerned components by density propagation. M consists of six components, respectively.

4. Experimental results

4.1 Seven movements

The proposed system was applied about seven basic movements used on movement education to be suitable for a development of a young children's. Seven basic movements are divided as locomotor movements and non-locomotor movements: the locomotor movement is changing a situation of the body in spaces as walking, running and hopping, and the non-locomotor movement is moving fixed the body as bending, stretching, balancing and turning. Walking is a movement moving one's weight from one leg to the leg. Jumping is a movement to jump with two on foot or two feet in the air, and get down with on foot. Hopping is the rhythmical movement landing with the same foot after one pushes up one foot with weight from the ground or floor. Bending is to make two close body parts approach one another through bending movement. Stretching is to stretch various parts of human body vertically or horizontally. Balancing can be conducted in a static state to support the center of body, or can carried out while one is moving. Turning is a subsequent movement turning the while body centered on the vertical or horizontal axis, while twisting is twist some part of human body. This paper was performed to five-year-old or six-year-old children in the neighborhood kindergarten. Fifteen young children repeated each movement from three to five times. A video camera was placed at the front of each young child.

4.2 Experimental environments and data

In the paper, video used to experiment carried out in Windows XP with a Pentium- IV 3.0 GHz CPU with 1GB of memory, and a program implemented in Microsoft VisualC++ 6.0. The video frames used in this experiment were acquired with a Sony DCR-PC330 at 15 frames per second (fps)

Table 2. Data number from test frames: (a) balancing, (b) bending, (c) hopping, (d) jumping, (e) stretching, (f) turning, (g) walking

movements	(a)	(b)	(c)	(d)	(e)	(f)	(g)	total
number of frames	208	223	163	186	190	204	196	8220

with a resolution of 320×240. The system was tested using a total 8220 frames (Table 2).

4.3 Experimental results

Fig.2 demonstrates 2-D locations and error distribution of each component for seven movements. Of all the sequences for movements, 10 frames are shown. In the first frame, although all the components showed large errors because of the initialization process, the error reduced over time.

Table 3 presents tracking rates of each

component in 8220 test frames and the proposed system achieved an average tracking rate of 97%. For each component, the tracking rates were as follows, head 97%, body 97%, left arm 98%, right arm 94%, left foot 98%, and right foot 97%.

Fig.3 shows the false tracking results of our experiments. As shown in Fig.3(a), since the young child's left arm and head overlap when stretching, they are detected as one group, in the image processing and thus the right arm should be separated from the head. In Fig. 3(b), a young child turns too fast; while an image component should have been tracked as a right arm, it was incorrectly tracked as body.

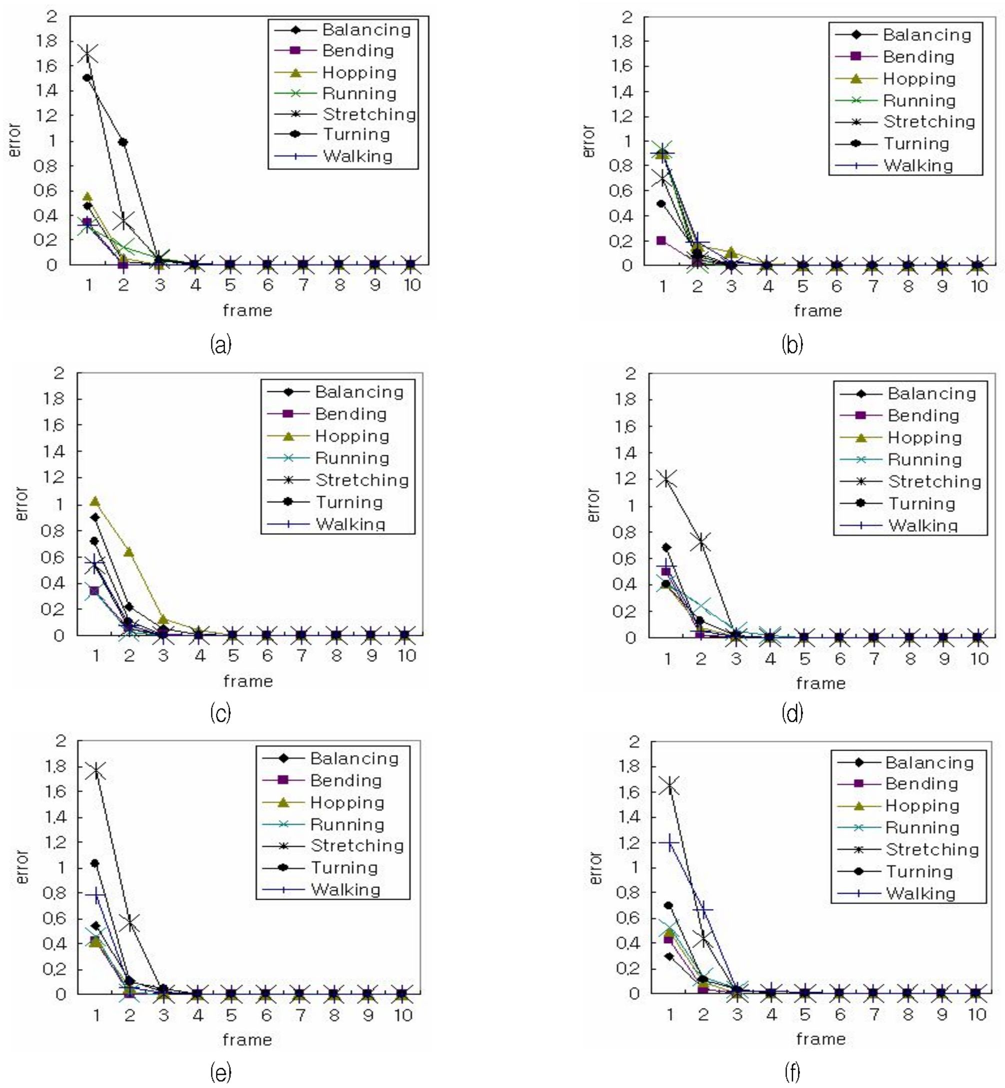
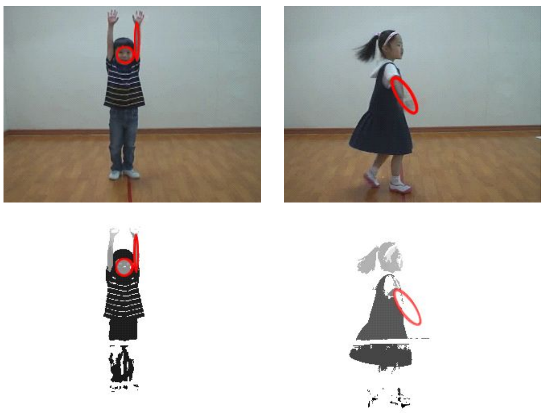


Fig.2. Error distribution of seven movements: (a) head, (b) torso, (c) left arm, (d) right arm, (e) left leg, and (f) right leg

Table 3. Tracking rates of each component in seven movements (%)

components movements	head	torso	left arm	right arm	left leg	right leg
balancing	97	97	97	95	99	96
bending	99	97	97	95	97	99
hopping	97	94	100	95	97	94
jumping	99	94	98	94	99	94
stretching	97	99	95	92	99	99
turning	98	97	99	92	97	98
walking	92	100	96	93	98	98
average	97	97	98	94	98	97



(a) (b)

Fig.3. False tracking examples: (a) error in wrong detection of a left arm and (b) error in fast turning. The top row shows the video frame and the bottom row the foreground. Ellipses indicate incorrectly detected components.

5. Conclusions

This paper presented a human body model suitable to tracking human body components for movement and proposed density propagation for this component-based human body model. Instead of using the original density propagation, for better tracking of a human body, we adopted density propagation to each body component that has geometric and appearance information and connection information. While each component was detected individually, the human body was tracked based on the component-based human body model. Hence, the proposed tracking method was applied to seven movements of young children and was verified by acquiring an average 97% of high tracking rate in Table 3. The proposed method is not only less sensitive to posture or lighting changes, but it also infers movements without markers; therefore, the proposed method makes tracking of a human body relatively straightforward.

Although not all the components are clearly detected, because of partial distortion arising from lighting conditions or overlapping, movements were detected as reliable using a component-based density propagation approach. For this reason, the proposed method has advantages of stable tracking of a broad range of movements.

In the future, a component-based tracking system proposed in this paper will be based on a silhouette, instead of being color based, and will be expanded to 3-D, using multiple cameras. In terms of the tracking method, future research will include use of the Kalman filter.

References

[1] O. Bernier and P. Cheung-Mon-Chan, "Real-time 3D articulated pose tracking using particle filtering and belief propagation on factor graphs", *BMVC*, pp. 27-36, Edinburgh: Sept. 2006.

[2] L. Campbel and A. Bobick, "Recognition of human body motion using phase space constraints", *MIT Media Lab. Perceptual Computing Group TR 309*, 1995.

[3] E. Clergue, M. Goldberg, N. Madrane and B. Merialdo, "Automatic face and gesture recognition for video indexing", IWAFGR, Zürich, pp.110-115, Switzerland: June 1995.

[4] M. Isard and A. Blake, "CONDENSATION-conditional density propagation for visual tracking", *Int. J. Computer Vision* 29(1), pp. 5-28, 1998.

[5] D. G. Lowe, "Robust model-based motion tracking through the integration of search and estimation", *Int. J. Computer Vision* 8(2), pp.113-122, 2004.

[6] A. Mittal, L. Zhao and L. S. Davis, "Human body pose estimation using silhouette shape analysis", AVSBS03, pp.263-270, 2003.

[7] B. Rosenhahn, U. G. Kersting, A. W. Smith, J. K. Gurney, T. Brox and R. Klette, "A system for marker-less human motion estimation", *LNCS* 3663, pp.230-237, 2005.

[8] R. Wang and W. K. Leow, "Human body posture refinement by nonparametric belief propagation", *ICIP*, pp.1217-1275, 2005.

[9] P. Alex, W. Christopher, H. David, D. Larry and H. Thanarat, "Real-time 3D Motion Capture", Second Workshop on Perceptual Interfaces, 1998.

[10] 이우람, 고한석, “모델 기반의 보행자 신체 추적 기법”, 전자공학회논문지, pp.601-699, 2007.

[11] O. D. Rishard, E. H. Peter, G. S. David, Pattern Recognition, ITC, 2006.

[12] 조경은, 조형제, "확률적 정규 문법 추론법에 의한 사람 몸동작 인식", 정보과학회논문지, pp.248-259, 2001.

[13] 박정주, 김주혜, 김명희, "피부색 정보 및 역 운동학을 이용한 인체 움직임 추정", 한국컴퓨터종합학술대회 2005 논문집, pp.820-822, 2005.

◐ 저 자 소 개 ◑

신 영 숙
1990년 2월 : 중앙대학교 전자계산학과(이학사)
1994년 2월 : 동국대학교 교과교육학과(전산교육전공)(교육학석사)
2001년 2월 : 연세대학교 인지과학과(전산학)(공학박사)
2003년 3월 ~ 현재 : 조선대학교 정보통신공학과 교수
관심분야 : 컴퓨터비전 및 패턴인식, 인공지능
E-mail : ysshin@chosun.ac.kr

차 은 미
2006년 2월 : 덕성여자대학교 인터넷정보학과 (공학사)
2006년 3월 ~ 현재 : 덕성여자대학교 대학원 교육영상매체학과 석사과정
관심분야 : 에듀테인먼트, 콘텐츠기획, 멀티미디어
E-mail : scott5@duksung.ac.kr

이 경 미
1993년 2월 : 덕성여자대학교 전산학과(이학사)
1996년 2월 : 연세대학교 전산과학과(이학석사)
2001년 12월 : 아이오와 주립대학교 전산학과(이학박사)
2003년 3월 ~ 현재 : 덕성여자대학교 컴퓨터공학부 교수
관심분야 : 내용기반 영상검색, 멀티미디어
E-mail : kmlee@duksung.ac.kr