

Kinect-based Gesture Password Recognition

Mohd Afizi Mohd Shukran, Mohd Suhaili Bin Ariffin

Faculty of Science and Defence Technology, Universiti Pertahanan Nasional Malaysia, Aras 6,
Bangunan Bistari, Kem Sungai Besi, 57000 Kuala Lumpur.

Abstract: Hand gesture password might be the most natural and intuitive way to communicate between people and machines, since it closely mimics how human interact with each other. Its intuitiveness and naturalness have spawned many applications in exploring large and complex data, computer games, virtual reality, health care, etc. Although the market for hand gesture password is huge, building a robust hand gesture recognition system remains a challenging problem for traditional vision-based approaches, which are greatly limited by the quality of the input from optical sensors. In this paper, we use their gesture in order to login or authenticate to the system. And then we introduce a novel method to create a gesture pattern that act as a password. This hand gesture recognition system performs robustly despite variations in hand orientation, scale or articulation. Moreover, it works well in uncontrolled environments with background clusters.

Key words: Password Recognition, Authentication, Gesture password

INTRODUCTION

The advent of relatively cheap image and depth sensors has spurred research in the field of object tracking and gesture recognition. One of the more popular devices used to do this type of research is Microsoft's Kinect, which has sensors that capture both rgb and depth data. Using similar data, researchers have developed algorithms that not only identify humans in a scene, but perform full body tracking; they can infer a person's skeletal structure in real-time, allowing for the recognition and classification of a set of full body actions (Wei Qu, A., 2007).

Kinect-based Gesture Password Recognition will require user to capture the gestures by point up and moving up the fingers or hand in front of the Kinect sensor without need to use a keyboard or mouse to key in password and the gesture itself as password. One of the known techniques to acquire gesture from a device or sensor is gesture recognition.

For Kinect-based games and applications, the most natural way to interact with the system is to recognize the gesture of a user. However, the Kinect camera is designed to detect the body pose and not hand gestures which required more precise algorithms (Michael H. Lin, 2004). This project will focus on testing efficient and accurate algorithms in detecting the fingers as a replacement for a password.

Gesture recognition is the mathematical interpretation of a human motion by a computing device. Gesture recognition, along with facial recognition, voice recognition, eye tracking and lip movement recognition are components of what developers refer to as a perceptual user interface (PUI). The goal of PUI is to enhance the efficiency and ease of use for the underlying logical design of a stored program, a design discipline known as usability. In personal computing, gestures are most often used for input commands.

Recognizing gestures as input allows computers to be more accessible for the physically-impaired and makes interaction more natural in a gaming or 3-D virtual world environment. Hand and body gestures can be amplified by a controller that contains accelerometers and gyroscopes to sense tilting, rotation and acceleration of movement -- or the computing device can be outfitted with a camera so that software in the device can recognize and interpret specific gestures (Jesús Martínez del Rincón, 2011). A wave of the hand, for instance, might terminate the program.



Fig. 2.1: Microsoft Kinect.

Corresponding Author: Mohd Afizi Mohd Shukran, Faculty of Science and Defence Technology, Universiti Pertahanan Nasional Malaysia, Aras 6, Bangunan Bistari, Kem Sungai Besi, 57000 Kuala Lumpur.
E-mail: afizi@upnm.edu.my

Using a Kinect, a user enters a combination of symbols by using a finger. The dimensions used to represent the password are the shape of each symbol, and the time taken when doing each of the traces. Using this representation, the user is forced to think of and memorize passwords in a more qualitative manner, and is also prevented from using bad habits such as storing passwords in text files. Kinect based password recognition system consists of using user finger to enter a pattern into a grid. This mechanism is far more to defend from key logger threat or brute force attack and man in the middle attack. Credit cards number, bank account's number, password or anything that key in with keyboard are being capture as typing text. The log will be used by third party to steal money, frauding and more control that cause loses to the victim (Leonid Raskin, 2011). In this mechanism, password is so called kept in the gesture pattern. There is no password disclosure from user. In other words, there is no typing text info are being capture by key logger. In the other hand, this proposed system will avoid user from forget their created password.

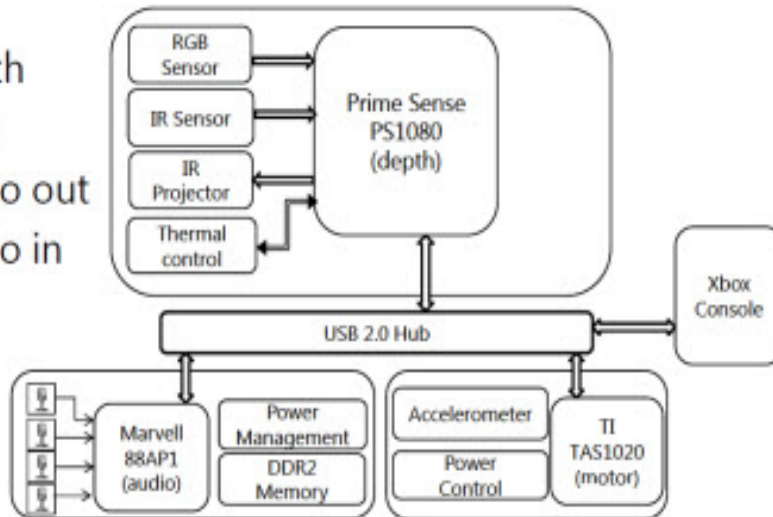
Gesture based password are widely used in mobile phone and tablet such as iPhone, iPad and Android. For Windows 8, Microsoft is a preparing a new way to log in to tablet PCs by letting users performs gestures on the screen instead of typing in letters and numbers. A user will choose a photo with some personal meaning to them, and create a sequence of taps, lines, and circles which must be performed in the right order to unlock the computer. The obvious question is whether such a system is as secure as typing a password on a keyboard. Given the kinds of simple passwords many users rely upon, the gesture-based system could well be more secure for numerous people. Microsoft acknowledges that smudges on the screen or recording devices could theoretically allow the gesture password to be compromised, but says the risk is very low.

MATERIALS AND METHODS

Kinect-based Gesture Password Recognition need user to use their gesture in order to login or authenticate to the system via Microsoft Kinect. Microsoft is a medium to capture the user's gesture in order to gain access to the system (Huazhong Ning, 2009). In this mechanism the password will be replaced by the gesture that have been created or captured.

Major Components & IC's

- USB 2.0 interface (Hub)
- Data
 - Depth
 - RGB
 - Audio out
 - Audio in
 - Tilt



8/3/2011

© 2011, Microsoft. All rights reserved.

9

Fig. 5.1: Microsoft Kinect's Components.

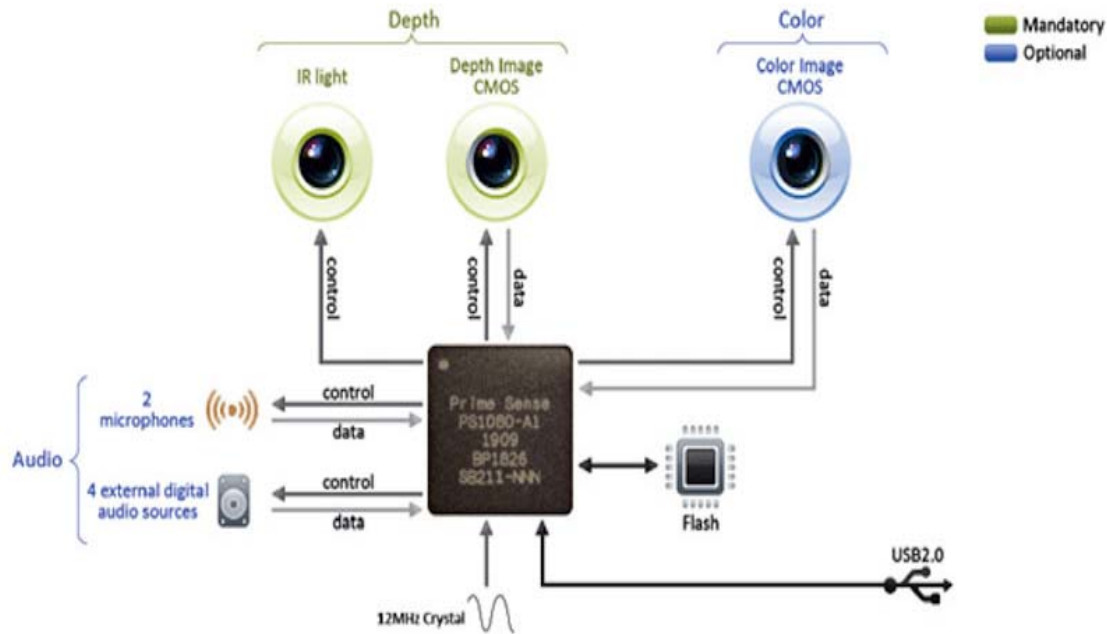


Fig. 5.2: Microsoft Kinect’s Camera.

Recognizing gestures as input allows computers to be more accessible for the physically-impaired and makes interaction more natural in a gaming or 3-D virtual world environment. Hand and body gestures can be amplified by a controller that contains accelerometers and gyroscopes to sense tilting, rotation and acceleration of movement -- or the computing device can be outfitted with a camera so that software in the device can recognize and interpret specific gestures (Olivier Bernier, 2009). A wave of the hand, for instance, might be the password.

With Kinect Based Gesture Password Recognition, system’s user will require to create a gesture pattern that act as a password. When user capture and draw a gesture pattern, the database will request for matching the gesture pattern for authenticate. A complete image of this flow can be referring to figure 5.3.

Gesture Types:

In computer interfaces, two types of gestures are distinguished: We consider online gestures, which can also be regarded as direct manipulations like scaling and rotating. In contrast, offline gestures are usually processed after the interaction is finished; e. g. a circle is drawn to activate a context menu (Xiaoqin Zhang, 2011).

Offline gestures: Those gestures that are processed after the user interaction with the object. An example is the gesture to activate a menu.

Online gestures: Direct manipulation gestures. They are used to scale or rotate a tangible object.

Gesture Implementation:

Gesture recognition is useful for processing information from humans which is not conveyed through speech or type. As well, there are various types of gestures which can be identified by computers (Shaobo Hou, 2007).

- Sign language recognition. Just as speech recognition can transcribe speech to text, certain types of gesture recognition software can transcribe the symbols represented through sign language into text.
- For socially assistive robotics. By using proper sensors (accelerometers and gyros) worn on the body of a patient and by reading the values from those sensors, robots can assist in patient rehabilitation. The best example can be stroke rehabilitation.
- Directional indication through pointing. Pointing has a very specific purpose in our society, to reference an object or location based on its position relative to ourselves. The use of gesture recognition to determine where a person is pointing is useful for identifying the context of statements or instructions. This application is of particular interest in the field of robotics.
- Control through facial gestures. Controlling a computer through facial gestures is a useful application of gesture recognition for users who may not physically be able to use a mouse or keyboard. Eye tracking in particular may be of use for controlling cursor motion or focusing on elements of a display.

- Alternative computer interfaces. Foregoing the traditional keyboard and mouse setup to interact with a computer, strong gesture recognition could allow users to accomplish frequent or common tasks using hand or face gestures to a camera.
- Immersive game technology. Gestures can be used to control interactions within video games to try and make the game player's experience more interactive or immersive.
- Virtual controllers. For systems where the act of finding or acquiring a physical controller could require too much time, gestures can be used as an alternative control mechanism. Controlling secondary devices in a car, or controlling a television set are examples of such usage.
- Affective computing. In affective computing, gesture recognition is used in the process of identifying emotional expression through computer systems.
- Remote control. Through the use of gesture recognition, "remote control with the wave of a hand" of various devices is possible. The signal must not only indicate the desired response, but also which device to be controlled.

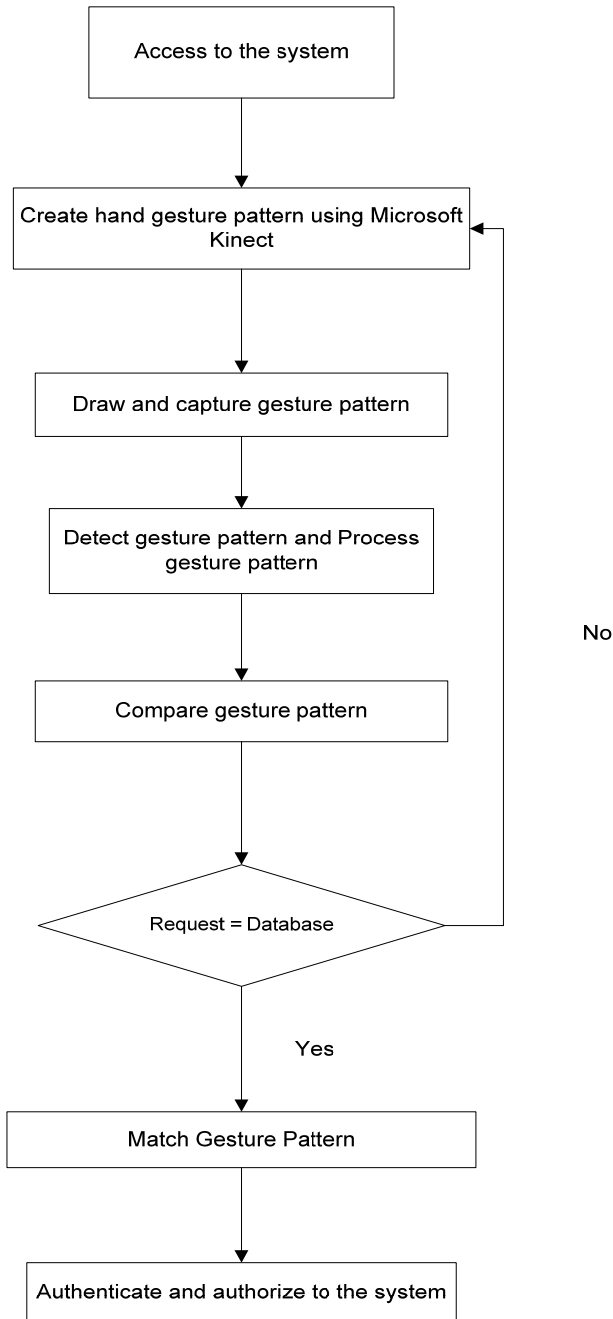
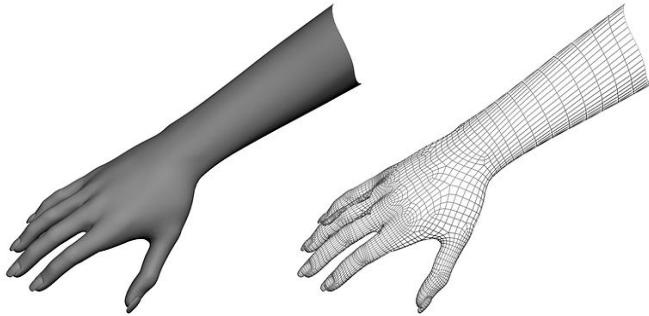


Fig. 5.3: System Flow.

Hand Gesture Algorithms:

Depending on the type of the input data, the approach for interpreting a gesture could be done in different ways. However, most of the techniques rely on key pointers represented in a 3D coordinate system. Based on the relative motion of these, the gesture can be detected with a high accuracy, depending of the quality of the input and the algorithm's approach. In order to interpret movements of the body, one has to classify them according to common properties and the message the movements may express.

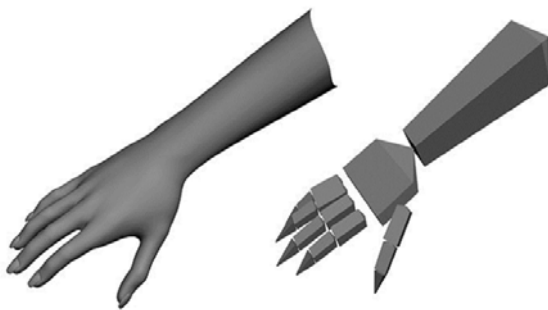
3D Model-Based Algorithms:



The 3D model approach can use volumetric or skeletal models, or even a combination of the two. Volumetric approaches have been heavily used in computer animation industry and for computer vision purposes. The models are generally created of complicated 3D surfaces, like NURBS or polygon meshes (Rong Zhu and Zhaoying Zhou, 2004).

The drawback of this method is that is very computational intensive, and systems for live analysis are still to be developed. For the moment, a more interesting approach would be to map simple primitive objects to the person's most important body parts (for example cylinders for the arms and neck, sphere for the head) and analyse the way these interact with each other. Furthermore, some abstract structures like super-quadrics and generalised cylinders may be even more suitable for approximating the body parts. Very exciting about this approach is that the parameters for these objects are quite simple. In order to better model the relation between these, we make use of constraints and hierarchies between our objects.

Skeletal-Based Algorithms:

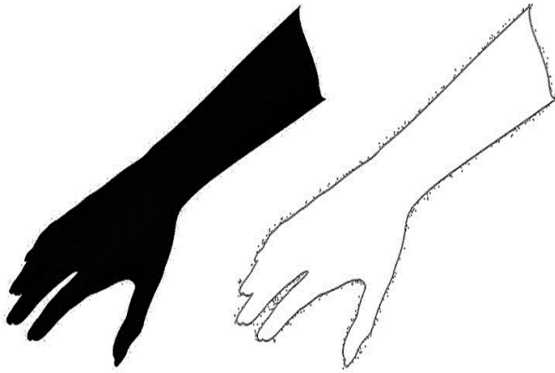


Instead of using intensive processing of the 3D models and dealing with a lot of parameters, one can just use a simplified version of joint angle parameters along with segment lengths. This is known as a skeletal representation of the body, where a virtual skeleton of the person is computed and parts of the body are mapped to certain segments (Ali Shahrokni, 2009). The analysis here is done using the position and orientation of these segments and the relation between each one of them(for example the angle between the joints and the relative position or orientation)

Advantages of using skeletal models:

- Algorithms are faster because only key parameters are analyzed.
- Pattern matching against a template database is possible
- Using key points allows the detection program to focus on the significant parts of the body

Appearance-Based Models:



These models don't use a spatial representation of the body anymore, because they derive the parameters directly from the images or videos using a template database. Some are based on the deformable 2D templates of the human parts of the body, particularly hands. Deformable templates are sets of points on the outline of an object, used as interpolation nodes for the object's outline approximation (Rong Zhu and Zhaoying Zhou, 2004). One of the simplest interpolation function is linear, which performs an average shape from point sets, point variability parameters and external deformaters. These template-based models are mostly used for hand-tracking, but could also be of use for simple gesture classification.

A second approach in gesture detecting using appearance-based models uses image sequences as gesture templates. Parameters for this method are either the images themselves, or certain features derived from these. Most of the time, only one (monoscopic) or two (stereoscopic) views are used.

Baum-Welch Algorithm:

Determining whether a sequence of observations fits the model requires us to solve a couple of probability problems. The first can be expressed as the probability of a certain true state I at time step t given the data:

$$Pr_t(i) = Pr(s_t = i|O)$$

If we can compute this marginal probability at every time step in a sequence, we can predict the hidden states and determine to what degree the data fit our model of a gesture. Recall that our model is a Markov chain, or a sequence of "random" events, so we're also interested in the transitions between states (Olivier Bernier, 2009). We can compute the marginal probabilities of the state being what it is (emission) and changing (transition) at every time step using a message-passing algorithm called the forward-backward algorithm where the messages are computed like:

$$\alpha_{t+1}(j) = \sum_{i=1}^N [\alpha_t(i) a_{ij}] b_j(o_{t+1})$$

$$\beta_{t+1}(j) = \sum_{i=1}^N a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)$$

The probability of the model being in a specific state I at time t and the probability of transition from state I to state j at time t can then be estimated:

$$\gamma_t(i) = \frac{\alpha_t(i) \beta_t(i)}{Pr(O|\lambda)}$$

$$\gamma_t(i, j) = \frac{\alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{Pr(O|\lambda)}$$

Ultimately, our model is going to encode these probabilities in the transition matrix A and the emission matrix B. After we've computed the marginal probabilities at time t as above, we can re-estimate these model parameters using a compelling, easy-to-compute iterative optimization algorithm that was shown by Baum and Welch, the Baum-Welch algorithm, which consists of two update equations, alpha and beta:

$$\hat{a}_{ij} = \frac{\sum_{t=1}^{T-1} \gamma_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)}$$

$$\hat{b}_j(o_k) = \frac{\sum_{t: O_t=o_k} \gamma_t(j)}{\sum_{t=1}^T \gamma_t(j)}$$

RESULTS AND DISCUSSION

Kinect-based gesture password recognition provide secured alternative for the traditional password. This mechanism is more to protect from brute force attack, man in the middle attack, dictionary attack, keylogger etc. Since Microsoft Kinect only available in certain applications and games, but this mechanism can be applied in the current system. The Microsoft Kinect itself still under research in detecting hand gesture and finger gesture, but the password gesture can provide more secure for future gaming system.

Kinect-based Gesture Password Recognition is another new approach to replacing password on log in mechanism. For decades, username and password are widely used for log in an account discovering major weakness that is passwords can often be stolen, accidentally revealed, or forgotten. For this reason, Internet business and many other transactions require a more stringent authentication process. No password or password less login mechanism will overcome a lot of password issue and bring much more benefits which are:

- When a key pass transfer across the network without using conventional text format make cracking tools and method are unable to define the password.
- Key logger is unable to capture typing text when there is no password key in involving.
- Using gesture identification, much more easily to memorize than word phrase. Forgotten password issue will be reduce
- Creating gesture module are replacing password text box will protect login system from brute force attack or dictionary attack.
- Besides, less or more, it will reduce system developer burden in term of security control.

Conclusion:

In conclusion, this paper has introduced hand gesture password in order to login or authenticate to the system. And then we introduce a novel method to create a gesture pattern that act as a password. This hand gesture recognition system performs robustly despite variations in hand orientation, scale or articulation. Moreover, it works well in uncontrolled environments with background clusters. The introduction of this novel method of authentication has produced several advantages such as Internet business and many other transactions require a more stringent authentication process. Besides that, hand gesture password would contribute not only in highly secure system but also can be developed more on gaming industry.

REFERENCES

- Ali Shahrokni, Tom Drummond, François Fleuret, Pascal Fua, 2009. "Classification-Based Probabilistic Modeling of Texture Transition for Fast Line Search Tracking and Delineation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(3): 570-576, DOI: 10.1109/TPAMI.2008.236.
- Huazhong Ning, Tony X. Han, Dirk B. Walther, Ming Liu, Thomas S. Huang, 2009. "Hierarchical Space-Time Model Enabling Efficient Search for Human Actions", *IEEE Transactions on Circuits and Systems for Video Technology*, 19(6): 808-820, DOI: 10.1109/TCSVT.2009.2017399.
- Jesús Martínez del Rincón, Dimitrios Makris, Carlos Orrite Uruñuela, Member and Jean-Christophe Nebel, 2011. "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*, 41(1): 26-37, DOI: 10.1109/TSMCB.2010.2044041.
- Leonid Raskin, Michael Rudzsky, Ehud Rivlin, 2011. "Dimensionality reduction using a Gaussian Process Annealed Particle Filter for tracking and classification of articulated body motions", *Computer Vision and Image Understanding*, 115(4): 503-519, DOI: 10.1016/j.cviu.2010.12.002.
- Michael H. Lin, Carlo Tomasi, 2004. "Surfaces with occlusions from layered stereo", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(8): 1073-1078, DOI: 10.1109/TPAMI.2004.54.
- Olivier Bernier, Pascal Cheung-Mon-Chan, Arnaud Bouguet, 2009. "Fast nonparametric belief propagation for real-time stereo articulated body tracking", *Computer Vision and Image Understanding*, 113(1): 29-47, January 2009, DOI: 10.1016/j.cviu.2008.07.001.

Rong Zhu and Zhaoying Zhou, 2004. "A real-time articulated human motion tracking using tri-axis inertial/magnetic sensors package", *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 12(2): 295-302, DOI: 10.1109/TNSRE.2004.827825.

Shaobo Hou, Aphrodite Galata, Fabrice Caillette, Neil A. Thacker, Paul A. Bromiley, 2007. "Real-time Body Tracking Using a Gaussian Process Latent Variable Model", In *Proceedings of the 11th IEEE International Conference on Computer Vision (ICCV 2007)*, pp: 1-8, 14-21. DOI: 10.1109/ICCV.2007.4408946.

Wei Qu, A., Dan Schonfeld, 2007. "Real-Time Decentralized Articulated Motion Analysis and Object Tracking From Videos", *IEEE Transactions on Image Processing*, 16(8): 2129-2138, DOI: 10.1109/TIP.2007.899619.

Xiaoqin Zhang, Xinghu Shi, Weiming Hu, Xi Li, Steve J. Maybank, 2011. "Visual tracking via dynamic tensor analysis with mean update", *Neurocomputing*, 74(17): 3277-3285, DOI: 10.1016/j.neucom.2011.05.006.