Face Spoof Detection by Motion Analysis on the Whole Video Frames

Heni Endah Utami
Graduate School of Informatics
Telkom University
Bandung, Indonesia
henien@student.telkomuniversity.ac.id

Hertog Nugroho
Electric Engineering Department
Bandung State of Polytechnic
Bandung, Indonesia
hertog@polban.ac.id

***Abstract*— Since the popularity of social media has increased, the spoofing attack is one of the many issues that occur in the face authentication system. Video replay attack is the most vulnerable type of spoofing attack, because distinguishing the face spoofing video from its original is very difficult. Most of the existing studies to detect face spoofing attack focused only on the face area and on static images. These approaches have weaknesses since in the real situation, the background texture and other body parts such as hair and shoulders have also been captured by the camera. In this paper, a face spoofing detection method using motion analysis in the whole video frame was discussed. To optimize the segmentation process between the object and the background area, the model-based segmentation was implemented so that all parts of the object's/user's body can be correctly separated from the background. The data used are 1200 videos from IDIAP Replay-Attack database and 500 videos from synthesized database for additional experimental data. The results of the experiment were that Half Total Error Rate (HTER) were 4% for IDIAP Replay-Attack data and 1% for synthesized data.**

Keywords—face spoof detection, image segmentation, motion analysis

I. INTRODUCTION

Authentication applications based on face identification are becoming popular and are increasingly developed since biometric data of the face (photos, videos) can be easily taken with available devices like cameras. In real application, a face is taken by a webcam on the computer or the laptop to be authenticated for the access control (lock and unlocks) of the computer. The webcam that was embedded in the computer allows a user to be in front of the camera (the center of view camera) with a certain distance so that the upper of the user's body was caught. Face authentication has also been widely used in identification system and access control in bankcard identification, online payment, gadget lock, security monitoring, etc.

However, this type of authentication is very vulnereble to spoofing attacks. In face authentication system, spoofing attack usually uses a printed photo in Fig. 1 (a), 3D facial mask in Fig. 1 (b), and video replay attack in Fig. 1 (c). Video replay attacks with display a video or photo are easier to launch than either printed photo attack or 3D mask attack and it is difficult to detect video replay attack since it is a replaying of the original live face and the high quality of this attack makes it similar with live faces. This is a type of attack which is most easily fool the camera-based face authentication system, particularly if the attacker uses the advanced recorder device to spoofing. It is a big threat to face authentication system, because this attack is very difficult to distinguish visually. So it is necessary to develop spoofing detection module focusing on video replay attack.

Fig. 1. (a) Print photo attack; (b) 3D mask attack; (c) Replay attack (Photo/Video).

II. RELATED WORK

Some researchers have conducted anti-spoofing for face authentication. Patel et al [1] addressed facial spoofing detection against replay attacks based on analysis of Moiré patterns, and evaluated the performance using different color channels. But visual evaluation on raw database showed that Moiré patterns only appeared on smartphone (low resolution) attack, and were not found in attack from high resolution device. Therefore, if the attacks come from high resolution devices, detecting Moiré patterns alone cannot detect face spoofing attack.

Chingovska and Anjos [2] attempted to use of the information about the identities of the enrolled clients. The result of this research was that the client-dependent approach outperforms the client-independent one with up to 50%. However, their work required face detection which required more computational cost. Furthermore, the information about the identities of the enrolled clients was taken from within the scope of the face area only. The feature information outside face area such as the body of user and background area was not utilized.

A new framework for face video spoofing detection using motion magnification was presented by Bharadwaj et al [3]. In their work, motion magnification was only used on image or normalized video by cropping the face region obtained from a commercial face recognition system. In the pre-processing step, they required eye detection to detect face area which was used

978-1-5386-3455-4/17/$31.00 ©2017 IEEE

to analyze. The same as the previous study [2], they did not use the feature information outside face area. Likewise Maatta [6] on spoofing detection from single image using micro-texture analysis, Tan [7] with presented face liveness detection from a single image with sparse low rank bilinear discriminative model and Pinto [11] with detection of face spoofing using visual rhythm, also only focused on face areas, and the video replay attack was produced by still images (static images).

Meanwhile, some works using information outside face area have also been proposed. Yan et al [8] proposed a method which included three scenic clues: non-rigid motion, face background consistency and image banding effect for accurate and efficient face liveness detection. However, similar with previous works [6, 7, and 11], they only utilized the video data produced from static images, and did not include dynamic images. Furthermore, they only evaluated the image with a complex background, because by using their method, a homogeneous background image was difficult to evaluate.

Pan et al [9] presented a real-time face liveness detection system using a monocular camera. Their approach was based on a combination of eye blinks and scene context. The authors used outside face clue of scene context called reference scene, which was similar to the background. Because their work was based on eye-blink and background matching, it was difficult to achieve if the attackers used video replay with eye-blink and had exactly similar background. Fathi et al [12] also performed similarity of background using Local Binary Pattern (LBP) to face spoofing detection. Using their own dataset, the accuracy of this approach was 98.9%. However, similar with [9], it was less effective if attackers used exactly similar background.

Komulainen et al [4] developed the spoof detection using fusion of motion and texture analysis. The motion features was taken froom the whole frame, while the textual features was extracted only from the square of face area. The consequence of this scheme was that the motion of the body of object outside the square area such as the shoulder and hair would be considered as background motion, while the textual features could also appear in the complex background condition.

III. The Proposed Method

Our goal is to build a tool that can detect video replay attack in indoor environments with complex background and lighting conditions from indoor and outdoor. We also tried to improve the accuracy from [4] by (a) implementing a model for segmentation process between the object/user and background area, so that the model is not only covered the face area, but also all parts of the object's/user's body and (b) using only motion analysis. The motion to be analyzed is defined as a changes of spatial features between frames, not only on specific features such as eye blink, mouth movement, etc.

The proposed method consists of two steps as is shown in Fig. 2. The first one is object/segmentation to separate the regions between the object and background area and the second one is classification. Segmentation was done by locating the object area using a model. That is why, as a preprocessing step, building a human body model should be performed. Each of the necessary steps, i.e: Building a model, body/background segmentation, and classification is described in the following subsections.

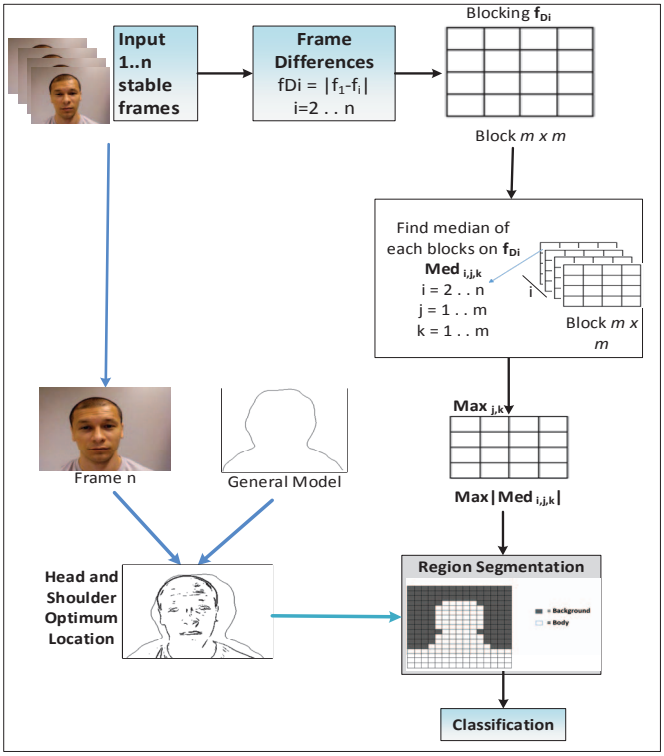

Fig. 2. Block diagram of the system

A. Building a Human Body Model

Motion analysis between the object (head and shoulder) and background region was the main topic. The goal of this process was to obtain a closed contour of the human upper body which could be used in segmentation process to separate object from background, especially in the image with complex background, and we call it 'model'.

Fig.3 shows the process to build a general model by combining the contours of selected objects with homogen background and the outermost boundary of the contours is extracted.

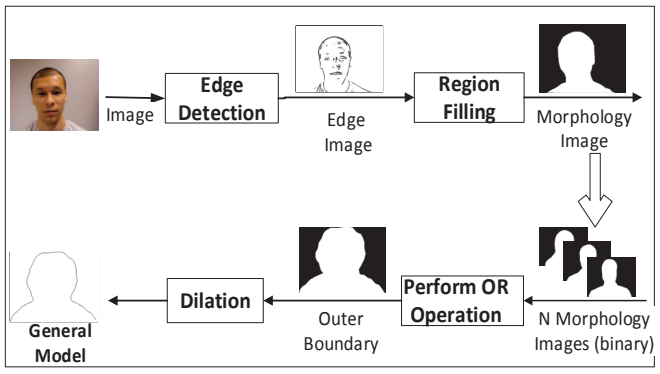

Fig. 3. Model building process.

B. Body/background Segmentation

In this work, locating the object area was a part of the region segmentation process. The essense of this part was to perform matching between the model obtained from the previous process and the input frames to extract the position of the user's upper body area in the frame.

The matching step to find the location of the object was done by finding the minimum difference value between model and frame *n* vertically (-*h* to *n*) and horizontally (-*w* to *m*). Suppose Mo was a model *m* x *n* and E_+ was a frame *m x n* with added *w* zero columns when Mo shifted *w* pixels horizontally, and added *h* zero rows when Mo shifted *h* pixels vertically. The Sum of Absolute Difference (SAD) between model (*Mo*) and frame *n* (*E*), could be calculated using equation (1):

$$SAD_{i,j} = \sum_{m\,x\,n} |E_+(m,n) - Mo(m+i, n+j)|, \quad (1)$$

$$-w < i < m,$$

$$-h < j < n$$

where *i,j* is the position of pixel at the most upper-left of the model. An optimum object (head and shoulder) location Ls_{opt}, is obtained by locating the minimum value of SAD vertically and horizontally :

$$Ls_{opt} = arg_{(i,j)} min\left(SAD_{i,j}\right), \quad (2)$$

$$-w < i < w,$$

$$-h < j < h.$$

The whole area of the object should be guaranteed to be located inside the model (not only attached to the model), including the cases when the object moved excessively. Therefore, after getting the optimal position of the object, the model should be enlarged. The segmentation process could be done only after the model had been enlarged.

On complex background image, the result of edge extraction as shown in the Fig. 4 (a). It seems that it is very difficult to distinguish between background and the object/user area. With the existing general model, the location/position of the object can be found by doing matching as shown in Fig.4 (c). Two regions are separated by the model in Fig. 4 (d): the area inside the model is regarded as the object and that outside the model is regarded as background.

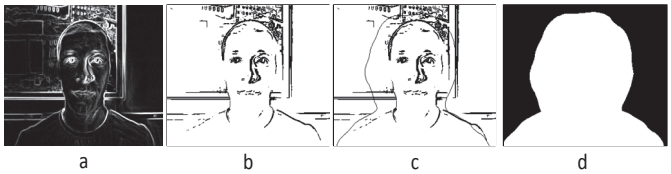

Fig. 4. (a) Edge detection on image with complex background; (b) binary image of image (a); (c) matching model with image (b); (d) the image of segmentation result

C. Classification

Before executing the experiment, stable frames needed to be found. Because, when the beginning process of capturing or recording user/client data, there are periods where the shadow appears or the device is not stable, for example, the camera device in an unsteady condition. These problems appeared at the beginning and end of the video and the stable frames were found between frames #61-#160. After selecting stable frames, and segmenting each frame into the foreground (object's upper body area) and the background, the next process was to collect motion features from the whole frames. Suppose we have *n* stable frames $f_i, i = 1..n$. To collect motion information from each frame, frame differencing fD_i between a first frame and the other frames was done using equation (3).

$$fD_i = |f_1 - f_i| \quad (3)$$

$$i = 2..n.$$

To get local motion information, each difference frame fD_i was divided into m x m blocks, and the median from each difference block was extracted.

$$M_{i,j,k} = Median(fD_{i,j,k}) \quad (4)$$

$$i = 2..n,$$

$$j = 1..m,$$

$$k = 1..m\ .$$

The median was taken because it could reduce the salt and paper noise on the image. After all difference blocks from all difference frames $M_{i,j,k}$ were collected, then the next process was to evaluate motion pattern in each difference block location through the whole frames. Since motion was indicated by $M_{i,j,k} \geq th$, the maximum value was taken from the whole difference frames $Max_{j,k}$ by equation (5).

$$Max_{j,k} = Max_{i=2}^{n} |M_{i,j,k}| \quad (5)$$

To represent motion information in that block area $B_{j,k}$, a threshold value *th* was used, and each block was compared with this value using equation (6).

$$B_{j,k} = \begin{cases} H & Max_{j,k} \geq th \\ L & else \end{cases} \quad (6)$$

Where *H* (High) was a symbol for the value 5-255 which means there are movement, while *L* (Low) is a symbol for the value 0-4 which means no movement. Therefore, the threshold *th* is 5.

The final step was to classify the video based on motion pattern in object and background area. Let $O_p \in B_{j,k}$, $p < m$ x m were a set of blocks belongs to object region and $BG_q \in B_{j,k}$ $q = m$ x $m - p$ were a set of blocks belong to background region. The region was classified using equation (7).

$$R = \begin{cases} H & \exists R = H \\ L & \forall R = L \end{cases} \quad (7)$$

where R can be O_p or BG_q. Detection for real video or spoof was done by using equation (8).

$$Result = \begin{cases} real & O_p = H \wedge B_q = L \\ spoof & else \end{cases} \quad (8)$$

Condition for real video was decided if there are movement (H) on the object area and no movement (L) on the background area.

IV. Experimental Result

A. Experimental Data

In experiment of this work, testing data of IDIAP replay attack database [13] and synthesized database were used.

The IDIAP Replay-Attack database [13] consisted of 200 video clips of the live/real client video, and 1000 video clips of photo and video replay attack for 50 subject with some conditions as shown in Fig. 4. The data were split into 3 sub-group, consisting of development data, training data and testing data. In this work, 360 video clips were used for threshold estimation of motion pattern, 360 video clips of training data were used for threshold of locating the object (model matching), and 480 video clips of testing data were used for experiment to report error figures.

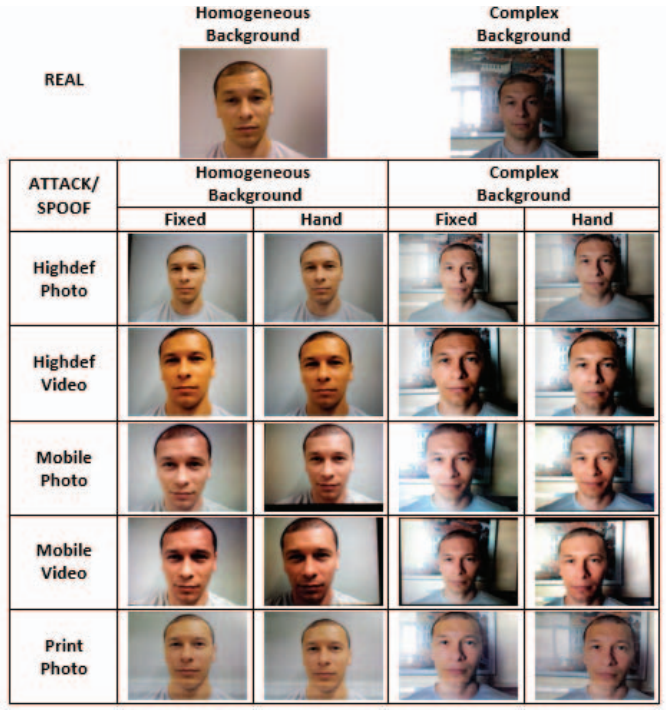

Fig. 5 IDIAP Replay-Attack Dataset

For additional experiment data, personal dataset were built, they consisted of 100 video clips of real clients and 400 video clips of replay attack for 50 subjects. Subjects consisted of 25 women and 25 men. Video recording was held on day and night with the lighting condition from indoor and outdoor of the post-graduate residency room in Telkom University and with the help of the light behind the camera to reduce the shadow. Real videos were produced by iMac webcam with software namely camera recorder with a resolution of 320 pixels by 240 pixels at 25 frames-per-second. Each client was recorded for 9 seconds under two different static background conditions (homogeneous and complex background) with a little movement in their upper body. The homogeneous background means a plain or uniform condition on the background, while the complex background means the background that has a composition of unrelated, or unlike elements or parts, varied or miscellaneous, such as: a window frame, portrait frame, tables, and other objects that have textures.

B. Experiment Measurement

A face spoof detection system has to deal with two kinds of conditions: either the face of the person recorded by the sensor is the real client (in which case, it is a real video), or the person is an impostor (in which case, it is a spoof video). Thus, the system may make two types of errors: a false acceptance, when the system accepts a spoof video, and a false rejection, when the system rejects a real video.

Suppose FA was the total number of false acceptances made by the system, FR was the total number of false rejections, NC was the number of client accesses, and NI was the number of impostor accesses. In order to be independent on the specific dataset distribution, the performance of the system was often measured in terms of rates of these two different errors, as follows:

$$FAR = \frac{FA}{NI}, \qquad FRR = \frac{FR}{NC} \quad (9)$$

In order to measure the performance of a spoofing detection system, the same measurement as the previous research[4] was used. It was the Half Total Error Rate, which combined the False Rejection Rate (FRR) and the False Acceptence Rate (FAR), where the probabilities were 0.5 each and was defined as [14]:

$$HTER = \frac{FAR+FRR}{2} \quad (10)$$

C. Experiment Analysis

TABLE 1 HTER OF VIDEO-BASED FACE SPOOFING DETECTION

Data	Number of Data	FA	FR	FAR	FRR	HTER
IDIAP Replay Attack	480 videos	2 of 400	6 of 80	0.5%	7.5%	4%
Synthesized Dataset	500 videos	0 of 400	2 of 100	0%	2%	1%

The results of the experiment are shown in Table 1 where video-based face spoofing detection in 480 video clips of

IDIAP Replay-Attack data with the HTER is 4%, and 500 video clips of synthesized dataset with the HTER is 1%. In IDIAP Replay-Attack testing of data experiments, 2 of 400 spoof videos are False Acceptance because 2 videos that should be spoof videos were detected as real videos. While 6 of 80 real videos were False Rejection because 6 videos were supposed to be real videos, detected as spoof videos. In this experiment personal dataset, 400 spoof videos were successfully detected as spoof videos and only 1 of 100 real videos were detected as spoof video. The results on Table 1 shows that the possibility of error in real video data is higher than spoof video data. This was because of problems that appeared in the background such as the shadow or background was not in a static condition in the real video. In that case a motion was detected and this made the video falsely classified as a spoof video.

TABLE 2 COMPARISON RESULT OF ERROR RATE USING HTER (%) BETWEEN THE PREVIOUS RESEARCHES WITH THE PROPOSED METHOD

Method	Dataset	Number of Data	HTER(%)
The previous research [4] by motion analysis	IDIAP Replay-Attack Testing data	Not available	11.2
The previous research [4] on by fusion	IDIAP Replay-Attack Testing data	Not available	5.11
Proposed method	IDIAP Replay-Attack Testing data	480 video clips	4
Proposed mothod	Synthesized data	500 video clips	1

Table 2 the comparison between the previous research [4] and the proposed method by model-based segmentation approaches to split the object and the background area. It shows that our results were better than the previous research [4]. Because in the previous method [4] on the motion analysis, a square area was used to segment the face from other, while motion in the whole frame was analyzed. The other area outside the square area (face area) was claimed as the background. So, the other part of the object such as hair and shoulder that have also motion information would be claimed as the motion of the background area, while in the real video, the background has a static condition. This shows the different results of motion analysis experiment between the previous research [4] and the proposed method. In the experiment for IDIAP Replay-Attack testing data, with model-based segmentation, the proposed method has a better result with the HTER was 4% than the previous research on motion analysis with the HTER was 11.2% because the model can cover the entire part of the object include hair and shoulder. The experimental results show that in the area of the object, information of motion not only appears in the face area but also on the entire object including the hair and shoulders. Moreover, the experiment result of the proposed method also outperformed the result of the previous research that was used fusion approach with the HTER was 5.11%. It was proved that with only using motion analysis, the system will have a low value of the error rate. So, the computational cost can be reduced while at least can still preserve the accuracy of the system.

V. CONCLUSIONS

This study proposed an approach for detecting video-based face spoofing by motion analysis, with model-based segmentation. The precondition in previous research [4] can be relaxed, because the user's face should not be inside the square area. By using model-based segmentation, motion on the whole part of the object including hair and shoulders could also be analyzed. It could be also implemented to ease the process of segmentation in complex background conditions. This study shows that the part of the object/user and the part of the background can be distinguished optimally.

In the video, if the background was still or no motion and any motions was detected in the object/user area, then the video was the real video. Whereas if it was detected that there was any motions both in the background and in the object's area or no motion at all in the whole frame, then the video was a spoof video. In the experiment, 480 video clips were used to test the data of IDIAP Replay-Attack database and 500 video clips of additional data were produced, with different types of attacks and different background conditions. The motion analysis was performed at 100 frames of the stable frames. The results of the experiments proved that by applying model-based segmentation to separate the object and background, the Half of Total Error Rate was able to achieve 4%, lower than the achievement of the previous work. This showed that by using motion information in the whole frame, video-based face spoofing could be detected.

This work still used the background in indoor with static conditions. It did not include the background in outdoor with dynamic conditions. Therefore, the future study needs to be able to detect face spoofing with dynamic background conditions including if shadow appears. For the segmentation process was also expected to be performed automatically.

REFERENCES

[1] K. Patel, H. Han, A. K. Jain, and G. Ott. Live Face Video vs. Spoof Face Video: Use of Moiré Patterns to Detect Replay Video Attacks. International Conference on Biometrics (ICB), 2015: 98-105.

[2] I. Chingovska, and A. Anjos. On the use of client identity information for face anti-spoofing. Journal Submission Special Issue On Biometric Anti-Spoofing, 2015.

[3] S. Bharadwaj, T. I. Dhamecha, M. Vatsa, and R. Singh. Computationally efficient face spoofing detection with motion magnification. In Proc. CVPR Workshops, pages 105–110, 2013.

[4] J. Komulainen, A. Hadid, M. PietikIainen, A. Anjos, and S. Marcel. Complementary countermeasures for detecting scenic face spoofing attacks. In International Conference of Biometric, 2013.

[5] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li. A face antispoofing database with diverse attacks. In Proc International Conference of Biometric, pages 26–31, 2012.

[6] J. Maatta, A. Hadid, and M. Pietikainen. Face spoofing detection from single images using micro-texture analysis. In Proc. International Journal Conference of Biometric, pages 1–7, 2011.

[7] X. Tan, Y. Li, J. Liu and L. Jiang. Face Liveness Detection from a Single Image with Sparse Low Rank Bilinear Discriminative Model. ECCV, 2010.

[8] J. Yan, Z. Zhang, Z. Lei, D. Yi, and S. Z.Li. Face liveness detection by exploring multiple scenic clues. 12th International Conference on Control, Automation, Robotics and Vision, (ICARCV2012), 2012.

[9] G. Pan, Lin S., Z. Wu, Y. Wang. Monocular camera-based face liveness detection by combining eye-blink and scene context. Telecommun Syst, (2011).47:215. https://doi.org/10.1007/s11235-010-9313-3.

[10] A. Pinto, W. R. Schwartz, H. Pedrini, and A. Rocha. Using Visual Rhythms for Detecting Video-based Facial Spoof Attacks. Transaction on Information Forensics and Security, 2015.

[11] A. S. Pinto, H. Pedrini, W. R. Schwartz, and A. Rocha. Video-Based face spoofing detection through Visual Rhythm Analysis. Conference on Graphics, Patterns and Images (SIBGRAPI), 2015.

[12] A. Fathi, and F. A. Mohammadi. Improving face recognition systems security using local binary patterns. 2015.

[13] I. Chingovska, A. Anjos, S. Marcel. On the Effectiveness of Local Binary pattern in Face Anti-spoofing. IEEE BIOSIG 2012.

[14] R Muthukrishn, M. Radha. Edge detection techniques for image segmentation. International Journal of Computer Science and Information Technology (IJCSIT) Vol 3, No 6 .Dec 2011.

[15] A. K. Jain., P. Flynn, A. A. Ross. Handbook of Biometrics, Springer, USA, 2008.

[16] S. Bengio, J. Mariethoz. A Statictical Significance Test for Person Authentication. ISCA 2004.

[17] H. Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W. T. Freeman. Eulerian video magnification for revealing subtle changes in the world. ACM Transactions on Graphics, 31(4), 2012.

[18] C.A. Glasbey, G.W. Horgan,. Image Analysis for Biological Science. Chapter 4. 1995.

[19] M. Sonka ,V. Hlavac, R. Boyle. Image Processing, Analysis, and Machine Vision. Third Edition. USA. 2008.