



HAL
open science

Multimedia Research Trends

Samarth Patel, Pamila Viswanathan

► **To cite this version:**

Samarth Patel, Pamila Viswanathan. Multimedia Research Trends. [Research Report] University of Alberta. 2021. hal-03183872

HAL Id: hal-03183872

<https://hal.science/hal-03183872v1>

Submitted on 29 Mar 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Trends in Multimedia Research

Samarth Patel

Dept. of Computing Science (Multimedia)
University of Alberta
Edmonton, Canada
samarth3@ualberta.ca

Pamila Viswanathan

Dept. of Computing Science (Multimedia)
University of Alberta
Edmonton, Canada
pviswana@ualberta.ca

Abstract—Over the years, tremendous progress has been achieved in many areas of multimedia research. Multimedia technology is widely applied everywhere and there exists a need for improvement due to the emerging technology and increase of data. In this paper we look into few of the many areas of research like Teleconferencing, Pose Estimation and Gesture Recognition, Medical Segmentation, Object Tracking etc.

1. Publication Review

A. Teleconferencing

1. Enhancing Videoconferencing Using Spatially Varying Sensing

In this paper [13] authors introduce a system for enhancing videoconferencing by using a compression system that incorporates multiple and moving foveae and combining information from them. Moreover, they also propose using spatially varying sensing as a pre-processor to JPEG. The authors propose using Variable Resolution transform to image compression, like the human visual system where in an area of interest has higher resolution and the peripheral scene has lower resolution. Moreover, Cartesian VR transform is used where the distortion factor is fixed, and the shape of the foveae is dependent entirely on its absolute location in the image and its position relative to other foveae. Moving foveae can be implemented on important feature such as mouth, eyes in a videoconferencing call to compress efficiently. Moreover, the model is extended to accommodate multiple foveae (cooperative and competitive foveae) and distortions. Furthermore, the concept of cooperative foveae is generalized as:

$$l_{actual} = \frac{1}{\sum_{j=0}^{N-1} \frac{1}{d_j^{power}}} \times \sum_{i=0}^{N-1} \frac{l_i}{d_i^{power}} \quad (1)$$

Where N is the number of foveae, power is the power with which weights corresponding to distances d_i are applied to l_i . The results observed after applying the technique as a pre-processor for JPEG are:

- JPEG with quality=1 after compression without VR: 3798 bytes (98.55% compression) and SNR RMS of 18.25

- JPEG enhanced by VR pre-processor : 377 bytes (98.56%) compression and the SNR rms is 47.23.

2. *Variable Resolution Teleconferencing* The author in this paper [11] describes the Variable Resolution (VR) based image compression and a few new interpolation methods are outlined with comparison between them. The VR transformation has high resolution in the fovea and low resolution in the periphery. It has two parameters – compression and alpha (- distortion in the edges). A pixel in polar coordinates are mapped to VR coordinates. The drawback in this is that the transformed images are not rectangular. A multi scaling factor is implied so that each will have maximum distance to the edge of the image and can be mapped to a rectangle. Sometimes there will be multiple foveas, in this case a location point is found in the two estimated points and then weights are added according to the distance from the point to the fovea. The compression performance is determined by compression ratio, time and quality. The effect of if low, then the compression is high. If is high then there is optimal compression. The new interpolation method depends on three aspects. One is generation of compression image, next is compressed final image last is the algorithm used to find the ‘nearest neighbour’. The nearest neighbour algorithm usually worked outwards to look for neighbouring pixels. This was modified to work on both the directions starting from the centre. Instead of computing each pixel in the compressed image, the pixel is computed as the average of all the pixels around it. For uncompressing they use the bi-linear method which uses the 4 nearest samples to find the pixel value. A comparison is done using a variety of combination of the method described. By adding more computation to the original method, the results can be improved. An intraframe difference coding to find the difference between two successive frames and to transmit only the difference. Also, no special boards are needed to perform this technique.

3. *Panoramic Video with Predictive Windows for Telepresence Applications* In this paper [14] the authors propose applying predictive Kalman filter to predict the fragment of a frame that is to be transmitted in a panoramic image taken using a CCD camera and conic mirror. The Kalman predictor is applied to a mouse modelled as: $mx = F - bx$.

where m is the mass, F is the external force applied and b is the coefficient of friction. Moreover, the Error covariance matrices Q_k for process noise and R_k for measurement noise are of the form: $Q_k = \text{diag}(q_{k^2}, q_{k^2}, q_{k^2})$ and $R_k = \text{diag}(r_{k^2}, r_{k^2})$. Here the values for q_k and r_k are determined by combination of techniques used in [16,17]. Moreover, the warping process takes images provided by the camera and based on point-of-view, field-of-view and aspect ratio, creates flat images that could be taken from a cylindrical view. Here the warping process uses a simple transformation to keep the processing time small. The prediction system predicts the likely position of the mouse in the next step and prepares the next fragment accordingly. In order to get efficient performance from the system the authors try to mitigate each of the three factors affecting response delays:

- Delay in capturing the next frame from the video stream with capture time varying between 0.033 seconds to 0.0143 seconds.
- Transmission time 0.15 seconds for grey scale image of 320 by 240 pixels.
- Client Side Processing: 0.05 seconds.

B. Pose Estimation and Gesture Recognition

1. Pose Recognition using Random Transform

The author proposes a novel method in [2] for human pose recognition using the Random Transform. There are other approaches used for pose recognition but they all need normalization and are also expensive. They propose a method without normalization and guarantees high resolution of leg and arm poses. Although Hough transform (HT) and Random Transform (RT) are used in various pattern recognition systems, due to some of the back draws in the HT they have used RT along with spatial maxima mapping. Their technique depends on two assumptions, i.e., any pose of the arm or leg can be replaced by the medial skeleton of that pose and the pose in RT space provides the pose classification information. The technique is divided into five steps. 1) The images are acquired and the background separation is done using the statistical background modelling approach. This provides a binary image of the arm and leg which are again processed to fill any holes present in the foreground. 2) These binary images undergo thinning to form a medial skeleton. 3) RT is used to detect the orientation of the skeleton. As in most cases the maxima are not localized or slightly diffused, they have included a threshold value to get the local maxima. The RT is then binarized as if RT value is greater or equal to the threshold then it is 1 else 0. An image dilation operation is performed on this so that there are no disconnected regions. The centroid location of each maxima is calculated and stored as f . 4) RT for a known pose B and unknown test pose T and its corresponding centroids are B and T . A one to one mapping is performed between the known and unknown pose. This is done for finding the maxima within a threshold limit. If a maxima is found then it is a 'hit' else is a 'miss'. The score is calculated from the ratio of hit to miss. 5) If the score is high then T is assigned

as the same pose as B .

2. Nose shape estimation and tracking for model-based coding

In this paper [3], the author proposes a nose shape detection method. This method limits search regions and provides a fix for shrinking effect of deformable templates using shape adaptive model. The nostril and nose side shapes detection and tracking is based on individual deformable templates.[4] is used for tracking the head movement. In this paper a two stage algorithm is developed for feature center location-global region growing and local region growing. In the former, a large threshold is set to detect the dark regions in the face while ignoring the rest. The face region is examined by selecting the center point in the head as a seed pixel. If the size of the growing region is determined within a range and the skin region is found. Later blobs (facial, top bottom) are found which are formed as four groups by performing vertical and horizontal clustering. In the later part, a small threshold is set to detect the facial organs. The feature center and its surrounding area is the local area of that feature. Here rather than the color they take the luminance to set a small growing threshold. The feature region of the organs are extracted from the binary image. After the features are detected, the shape of the eye is found using the Hough Transform (iris position and size). For the nose, the nostrils shape is detected using a geometric template(twisted pair curve). The width is determined by finding the distance of both the curves, the orientation by eigen vectors and the shape is controlled by a parameter which ranges between 1 to 10. The nose side is estimated using a vertical parabola. Finally a 3D wireframe is matched on the face to track the facial movements.

3. A Multisensor Technique for Gesture Recognition Through Intelligent Skeletal Pose Analysis

In the paper[19] a novel technique for gesture recognition is proposed using multisensor based intelligent skeletal pose analysis. One of the primary problems tackled by the authors is estimation of hand poses when then hand is not directly facing camera, or parts of it are occluded such as in the case of a pinch gesture where some fingers are occluded by others which could prevent accurate gesture recognition.

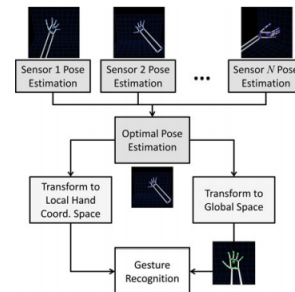


Figure 1. Overview of system

In the proposed method the optimal pose is estimated using a trained support vector machine, which is built once with a training set which uses a feature vector composed

of subsets of each sensor's output. Subsequently each sensor's pose output is converted into global coordinate system to represent the poses of all fingers in a single unified space, while also keeping the positions of all fingers in local coordinate system, finally both the local and global hand pose information are input into the gesture recognition model. Compared to single sensor-based approach with a mean pose estimation error of 15.48 the proposed method has mean pose estimation of 10.60. This method is 90.8% optimal in terms of reducing the mean pose estimation error. Moreover, for correctly classified instances the median difference between pose estimations was 11.44mm and for incorrectly classified instances the median difference was 4.47mm, indicating that in cases of large disagreements between two sensors, the proposed approach often made the correct decision.

C. Object Tracking

1. Eye Tracking and Animation for MPEG-4 Coding

This paper [5] discusses on the eye detection and tracking in a model based approach. The eye movement is detected by using Hough Transform and deformable template matching similar to the feature detection in [6]. In the proposed algorithm the face region is detected and that the upper right and upper left coarse regions can be used to detect the eyes. The iris of the eyes is detected using the Hough Transform. The intensity image obtained is convolved with a Sobel filter and all the edges with lower magnitude is eliminated. The exact center of the iris is got by the Hough Transform. The eye model described in this paper has parameters shown in the figure.

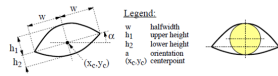


Figure 2. Deformable template model of the eye

The potential energy is created by matching deformable templates using two images. The template is positioned approximately relative to the iris. The h_1 and h_2 parameters are then approximated. h_1 is given more importance as the upper eyelid has more movements. The horizontal integral projection along with the gradient projections in both the directions, the eyelid position is estimated. The eyelid are then matched to the deformable template. In the tracking of the eye movements the position and size of the eye are restricted to the previous frame also the matched template will be used to initialize the next template. For the animation of eyes, 11 feature vertices are used in each eye to match the corresponding feature points on the template. The iris animation is done by matching the center of the iris model with the center of the eye image. The eyelid animations are also done by matching the feature vertices of the model to that of the image. Its displacement consist of feature motion vectors and non-feature motion vector are found by interpolation the feature vectors that are around them.

Based on these vertices a deformed eye model is built for each frame. Finally, the eye movements of the model are synthesized by texture mapping the original image.

2. Integrating active face tracking with model based coding

This paper [8] proposes a system for tracking a face with an active camera using model based coding approach. The proposed algorithm can be divided into 4 major steps. 1) Initially a background compensation is performed to estimate the relationship of the pixels in a 3D point in images taken at different camera orientations. In this process only the camera coordinates and the image plane have changes in direction. The pixel in the previous image could be calculated by finding the pixel position in the current image. 2) This paper points the necessity to get a head silhouette from a set of frames. The difference between the current and previous frames are found and the absolute value is taken as a threshold. Morphological filtering is applied to remove the noise and multiple frames are integrated using frame motion fusion algorithm called motion mask (MM). To fuse the consecutive masks a spatiotemporal filter is used. It is then smoothed more by the spatial median filter. Threshold is set as 1 and is applied to the fused image. Further a morphological filter is applied to smoothen the image and by using the gradient operator the edges are detected. 3) The facial features are then detected using deformable template matching and exploit color information. The eye detection has been explained in [5]. The mouth detection is done by fining the region of interest. The approximation of the lip is got from the color information which is then localized using deformable templates. 4) The facial model adaptation is done from two image views. As the shape of the head is known, the eyes and the mouth can determine the face position. A 3D wireframe model is fitted into the facial information extracted using an energy minimization procedure.

D. Medical Segmentation

1. Automatic Segmentation of Spinal Cord MRI Using

Symmetric Boundary Tracing In this paper [15], the authors propose a fully automated active contour tracing algorithm for extraction of spinal cord from MRI Here the image is first Unskewed as a precursor to segmentation of other targets. After drawing a circle at the centre of mass of the spinal cord region, Using the Bhattacharya coefficient between the intensity histograms of sectors, the two sectors obtained by drawing a diameter in the circle are matched. After the image is Unskewed, an evolved active trace algorithm is applied to segment the muscle region. The failure of correct tracing caused by uniform search regions of each active trace point along its normal direction is solved by using temporal constraint in order to introduce variable search space for each active trace point along its orthogonal direction. The overall complexity is $O(nm^3)$. Compared to existing methods which extract only the spinal cord, this approach segments the neighbouring regions of interest providing better visual guidance for the Intraspinal micro stimulation (ISMS) implant.

2. Gradient Vector Flow based Active Shape Model for Lung Field Segmentation in Chest Radiographs

The paper [7] explains the Gradient Vector Flow (GVF) with Active Shape Model (ASM) – a high level method to accurately segment the lung field from chest radiographs. This paper explains three types of previous related work. 1) PDM generation – There are two models in this approach, shape model and profile model. The shape model is formed from a set of training images. The principal component analysis is done to approximate the training dataset. In the profile model, the PCA is done to get the grey level variation parameters. 2) ASM based segmentation – a PDM is performed to estimate the contours that are near to the lung field in the test image. Matches are found between the control points found above to the target points. A displacement is done to calculate the shape of the model. 3) GVF-ASM Segmentation – Along with the ASM segmentation, a GVF calculator to find the edge map and GVF-based points evaluation to add both magnitude and direction. In the proposed method, it overcomes the blurring in the edges of the contour field as the magnitude and gradient vectors changes smoothly by proposing a new points evaluation equation:

$$C = C' + w \cdot \text{sgn}(g(C')) \cdot e^{-g(C')} \quad (2)$$

This helps to decrease the smoothness in the contour edges and gives point to strong edges. In comparison with the GVF-ASM segmentation, the proposed method when there is a change in parameters is still accurate in the edges and the other is sensitive to parameter selection. It is also proven that the proposed method is more stable and accurate whereas the GVF-ASM has a strong ringing effect and moves away from the edges.

Other Fields

1. Modelling Fish-Eye Lens

In this paper [1] the author proposes a variable resolution transformation based on the characteristics of fish-eyes without using any sensors. There are two methods described for modelling fish-eye lens. Firstly, the introduced Fish-Eye transformation (FET) generates a high resolution in the fovea and a low resolution in the periphery. It also points out the anisotropic distortion present in the peripheral regions of the image. Given various distortion parameter, different fish-eye transformation images is generated. Secondly, the Polynomial Fish-Eye Transformation (PFET). In this paper, PFET modelling of data is done by fitting a set of observation to a model which is a polynomial. It is similar to FET except that it includes the radial distance for undistorted images. The optical center of the distortion, called the focus of distortion (FD) is found by the average minimum curvature of the intersection of the lines. Using FD, the measure of the average of the distortion in all four directions are obtained. Using the polynomial function and a set of data pairs and the Gauss-Jordan elimination method is used for finding the average distortion. The degree of the

polynomial uses the variance of the model and the fitted data. For FET as the unknown coefficients are nonlinear with the equation, instead of Gauss-Jordan, Newton's method is used. The PFET shows a better approximation but FET can easily modify distortion by and can restore the original image using inverse function but not in PFET

2. Stereo Matching Using Random Walks

This paper [9] presents a novel approach for stereo matching using random walks (RW). It is a two-step process where in the first step a set of reliable matching pixels are determined using RW and in the second, using these reliable points the disparity of unreliable pixels are determined. The Random walks is a graph representation of an image where the probability of a walker from an unlabelled node to a labelled node is calculated which is equal to the Dirichlet problem. The prior matrices are built according to the illumination in the image and the accuracy of the resulting disparity maps are performed by the Laplacian matrix. In this process, the Laplacian matrix and the prior matrix are built first. If the brightness is low, then the similarity difference between the pixels will be small, hence an average luminance value is used for the measurement of brightness. A left-right checking is done to the set of reliable pixels. Its process is similar to [10] which results in a set of matched pixels (M) and the rest as unreliable set (U). The textureless regions have similar pixels around its neighbours so a low disparity is obtained, which is incorrect. After the left-right checking, every pixel goes through a disparity validation test. Finally the disparity of unreliable pixels are computed by solving the Dirichlet problem. The drawback in this is that it works only in YCbCr space, so it is limited to color images.

3. A Framework for Adaptive Training and Games in Virtual Reality Rehabilitation Environments

In the paper [12] the authors propose a low-cost, low-maintenance, high availability, system for training people with disabilities in the use of powered wheelchair using virtual reality, moreover they also propose an approach using Bayesian networks to enable adaptive training in virtual rehabilitation environment. The system consists of training program that adaptively changes difficulty according to the performance of the patient thus taking into account the range of disabilities and residual capabilities that the patient may have, and also an XML scene node-based framework as a 3D-modelling plug-in tool for clinician to design and alter position, scale, rotation of objects in the training space to customize the environment to patient need based on their assessments. Moreover, the adaptive system modelled using Bayesian networks can be easily extended to include large number of latent and measured variable with complex interactions between them. During tests, the system was found to effective with a mean time of 81.4 seconds for experimental group that was trained using this system before their attempt to complete a short obstacle course using a powered wheelchair compared to a mean time of 104.5 seconds for the group that did not receive any training with this system.

4. Interactive Multimedia for Adaptive Online Education

In this paper the authors propose an Interactive Multimedia program based on a novel CROME (Computer Reinforced Online Multimedia Education) framework which integrates all main component of education i.e learning, teaching, and testing. The proposed system is capable of adaptively altering the difficulty of questions during testing. Moreover, also consists of custom-designed modules for tracking precise concepts, Computer assisted question item-authoring. The proposed incorporates rich interactive elements tailored to the content for instance students are presented chemical equations with 3D objects that they can interact with, moreover it is also capable of drag and drop interaction for both 2D and 3D objects, It has language items that combine acoustic and visual effects to help students relate semantics to spellings and pronunciations. The proposed system is implemented using Java 2D/3D applets, JavaScript, PHP and MySQL following the principles of portability, Reusability, Interoperability, and Scalability.

5. *Variable-resolution character thinning* The paper introduces a variable resolution character thinning approach modelled after the working of human eye, in the sense that the foveae is higher resolution compared to periphery. The proposed algorithm is parallel and uses a 9x9 window with a high resolution 3x3 centre and low-resolution periphery. The resolution is reduced by a ratio of 9:2 in the periphery. Here all character pixels are first examined by Guo's algorithm A1 [21], while also comparing each character pixel with a 3x3 template. If certain constraints are satisfied, then another set of 9x9 templates are used. Here the centre part is a regular 3x3 template, while each of the eight peripheral parts also covers a 3x3 region, but at reduced resolution. Each peripheral region generates two binary values α and β , where

$$\alpha = \begin{cases} 0 & \text{if there are } \leq 5 \text{ character pixels in} \\ & \text{this region,} \\ 1 & \text{otherwise.} \end{cases}$$

β is used to record some special pattern of pixel arrangement in the 3x3 region. The results indicate that the proposed approach of variable resolution character thinning performs better in accurately thinning the given character by preserving the straight lines in the characters compared to other approaches such as Holt's method and Guo's method.

References

- [1] A. Basu and S. Licardie, "Modeling fish-eye lenses," Proceedings of 1993 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '93), Yokohama, Japan, 1993, pp. 1822-1828 vol.3, doi: 10.1109/IROS.1993.583883.
- [2] M. Singh, M. Mandal and A. Basu, "Pose recognition using the Radon transform," 48th Midwest Symposium on Circuits and Systems, 2005., Covington, KY, USA, 2005, pp. 1091-1094 Vol. 2, doi: 10.1109/MWSCAS.2005.1594295.
- [3] Lijun Yin and A. Basu, "Nose shape estimation and tracking for model-based coding," 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.01CH37221), Salt Lake City, UT, USA, 2001, pp. 1477-1480 vol.3, doi: 10.1109/ICASSP.2001.941210.
- [4] L. Yin and A. Basu. Integrating active face tracking with model based coding. Pattern Recognition Letters, 20 (6), 1999
- [5] S. Bernogger, Lujun Yin, A. Basu and A. Pinz, "Eye tracking and animation for MPEG-4 coding," Proceedings. Fourteenth International Conference on Pattern Recognition (Cat. No.98EX170), Brisbane, QLD, Australia, 1998, pp. 1281-1284 vol.2, doi: 10.1109/ICPR.1998.711935.
- [6] G. Chow and X. Li. Towards a System for Automatic Facial Feature Detection. Pattern Recognition, 26(12):1739-1755, Dec. 1993.
- [7] Xu, Tao Mandal, Mrinal Long, Richard Basu, Anup. (2009). Gradient Vector Flow based Active Shape Model for Lung Field Segmentation in Chest Radiographs. Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference. 2009. 3561-4. 10.1109/IEMBS.2009.5334886.
- [8] Lijun Yin, Anup Basu, "Integrating active face tracking with model based coding," Pattern Recognition Letters, Volume 20, Issue 6, 1999, Pages 651-657, ISSN 0167-8655
- [9] R. Shen, I. Cheng, X. Li and A. Basu, "Stereo matching using random walks," 2008 19th International Conference on Pattern Recognition, Tampa, FL, USA, 2008, pp. 1-4, doi: 10.1109/ICPR.2008.4761512.
- [10] C. Lei, J. Selzer, and Y.-H. Yang. Region-tree based stereo using dynamic programming optimization. In Proceedings of CVPR'06, pages 2378-238, 2006.
- [11] A. Basu, A. Sullivan and K. Wiebe, "Variable resolution teleconferencing," Proceedings of IEEE Systems Man and Cybernetics Conference - SMC, Le Touquet, France, 1993, pp. 170-175 vol.4, doi: 10.1109/ICSMC.1993.390703.
- [12.] Rossol, Nathaniel Cheng, Irene Bischof, Walter Basu, Anup. (2011). A framework for adaptive training and games in virtual reality rehabilitation environments. Proceedings of VRCAI 2011: ACM SIGGRAPH Conference on Virtual-Reality Continuum and its Applications to Industry. 10.1145/2087756.2087810.

[13.] A. Basu and K. Wiebe, "Enhancing videoconferencing using spatially varying sensing," in *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 28, no. 2, pp. 137-148, March 1998, doi: 10.1109/3468.661143.

[14.] J. Baldwin, A. Basu and H. Zhang, "Panoramic video with predictive windows for telepresence applications," *Proceedings 1999 IEEE International Conference on Robotics and Automation (Cat. No.99CH36288C)*, Detroit, MI, USA, 1999, pp. 1922-1927 vol.3, doi: 10.1109/ROBOT.1999.770389.

[15.] Mukherjee DP, Cheng I, Ray N, Mushahwar V, Lebel M, Basu A. Automatic segmentation of spinal cord MRI using symmetric boundary tracing. *IEEE Trans Inf Technol Biomed.* 2010 Sep;14(5):1275-8. doi: 10.1109/TITB.2010.2052060. Epub 2010 Jun 7. PMID: 20529752; PMCID: PMC3038784.

[16.] Satoru Emura and Susumu Tachi. Compensation of time lag between actual and virtual spaces by multi-sensor integration. In *Proceedings of the 1994 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, pages 463-469, October 1994.

[17.] Jiandong Liang, Chris Shaw, and Mark Green. On temporal-spatial realism in the virtual reality environment. In *UIST'91*, 1991.

[18.] I. Cheng, A. Basu and R. Goebel, "Interactive Multimedia for Adaptive Online Education," in *IEEE MultiMedia*, vol. 16, no. 1, pp. 16-25, Jan.-March 2009, doi: 10.1109/MMUL.2009.11.

[19.] N. Rossol, I. Cheng and A. Basu, "A Multisensor Technique for Gesture Recognition Through Intelligent Skeletal Pose Analysis," in *IEEE Transactions on Human-Machine Systems*, vol. 46, no. 3, pp. 350-359, June 2016, doi: 10.1109/THMS.2015.2467212.

[20.] Li, X. and A. Basu. "Variable-resolution character thinning." *Pattern Recognit. Lett.* 12 (1991): 241-248.

[21.] Guo, Z. and R. Hall (1989). Parallel thinning with two sub iteration algorithms. *Comm. ACM* 32(3), 359-373