



**HAL**  
open science

## Computer Vision Papers Literature Review

Sravanti Chinta, Rajat Bothra Jain

► **To cite this version:**

| Sravanti Chinta, Rajat Bothra Jain. Computer Vision Papers Literature Review. 2022. <hal-03606640>

**HAL Id: hal-03606640**

**<https://hal.science/hal-03606640v1>**

Preprint submitted on 12 Mar 2022

**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.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

# Literature Review

Rajat Bothra Jain  
Department of Computing Science  
University of Alberta  
Edmonton, Canada  
bothraja@ualberta.ca

Sravanti Chinta  
Department of Computing Science  
University of Alberta  
Edmonton, Canada  
sravanti@ualberta.ca

**Abstract**—The literature review presented below on Image Compression, Transmission of 3D data over wireless networks and tracking of objects is the in depth study of Research Papers done in Multimedia lab. Most of the papers presented in this literature review have tackled the problems present in the conventional system and offered an optimal and practical solution.

## I. GAUSSIAN AND LAPLACIAN OF GAUSSIAN WEIGHTING FUNCTIONS FOR ROBUST FEATURE BASED TRACKING

### A. Authors

Meghna Singh, Mrinal K. Mandal , Anup Basu

### B. Summary

This paper addresses the effects of noise on the performance of the standard KLT algorithm and proposes two solutions - Laplacian of Gaussian (LoG) and Gaussian weighting function [1] to increase immunity of the tracking algorithm to effects such as poor CCD exposure and compression noise. The KLT algorithm is limited by its use of averaging as a weighting function over the feature window. KLT algorithm fails to track the object even at low levels of noise, while the weighted tracking algorithm continues to track until 25% noise. Fig. 1 shows the affect of noise on performance tracking algorithm.

### C. Analysis

Irrespective of the input test sequences, both the proposed weighting functions, Gaussian weighting function and LoG weighting function, consistently perform better tracking in a random noise environment than the KLT average weighting function. Real world tracking sequences are expected to be noisy, and therefore are ideal for application of these weighting functions. The LoG is suitable for images with well-defined and sharp image corners like the synthetic test images. Real images where the edges may not be as sharply defined, the Gaussian weighting function performs better.

## II. INTEGRATING ACTIVE FACE TRACKING WITH MODEL BASED CODING

### A. Authors

Lijun Yin, Anup Basu

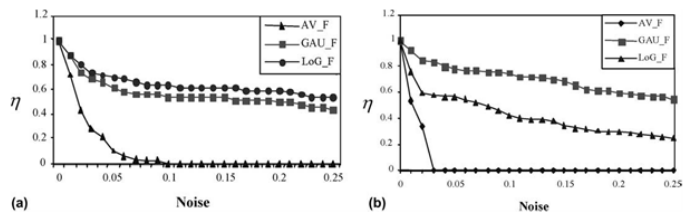


Fig. 1. (a) Performance of weighting functions with synthetic car sequence, 50 features have been initialized and are being tracked. (b) Performance of weighting functions with real image car sequence, 50 features have been initialized and are being tracked.

### B. Summary

This paper discusses the tracking of a talking face with an active camera whereas most previous work done focuses on the detection on a still camera. The detection and tracking of the active talking face is a fundamental step towards realizing an application of MPEG4 in real situations. First, the background compensation in successive frames is done, and [2] the motion-energy tracking approach is used coupled with a morphological filter to reduce the noise. Third, the facial features are detected using Hough Transform and deformable template coupled with color information. Finally, a wireframe model is adapted to the extracted face using a coarse-to-fine adaptation algorithm.

### C. Analysis

Proposed a system for tracking a face observed with an active camera for an arbitrary background. Camera mounted on an active platform (pan/tilt) is used to take an active video sequence, which shows a talking person with an unconstrained background Initial results show that this approach is feasible in practical applications.

## III. A FRAMEWORK FOR ADAPTIVE TRAINING AND GAMES IN VIRTUAL REALITY REHABILITATION ENVIRONMENTS

### A. Authors

Nathaniel Rossol, Irene Cheng, Walter F. Bischof, Anup Basu

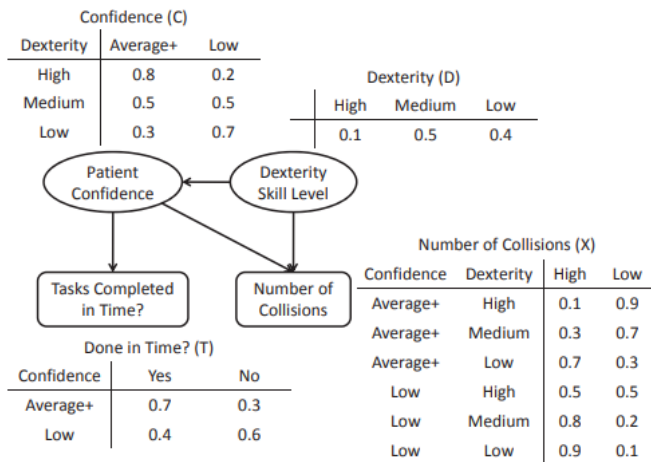


Fig. 2. A sample Bayesian network topology

### B. Summary

[3] Bayesian networks as a way to measure student performance in online web-based multimedia educational games but have not been used in Virtual Reality Rehabilitation. Fig. 2 shows the Bayesian network topology. This paper proposed a novel framework based on Bayesian networks for self-adjusting adaptive training in virtual rehabilitation environments. The system offers standard training modes and interactive virtual reality games for increased patient engagement. In addition, the system provides a framework to allow clinicians to design, build, and customize interactive indoor rehabilitation training environments that can dynamically respond to user actions in a meaningful way.

### C. Analysis

Participants who used the proposed virtual reality wheelchair training system completed the real world obstacle course on average faster, with a mean time of 81.5 seconds for the experimental group compared to 104.5 seconds for the control group .

## IV. INTERACTIVE MULTIMEDIA FOR ADAPTIVE ONLINE EDUCATION

### A. Authors

Irene Cheng, Anup Basu and Randy Goebel

### B. Summary

Through computer reinforcement, this paper addresses a vision of providing publicly accessible education anywhere, at any time, and to anyone. Online Multimedia Education [4](CROME) combines the primary components of education, such as learning, teaching, and testing, as well as adaptive testing and student modelling. Online tutoring addresses issues such as time difference, curriculum diversity, and potential language and cultural barriers. The use of digital multimedia and e learning is a more widely adopted and effective approach to promoting global education. We interpret a molecular

structure as a graph, where nodes are atoms and edges are bonds, to evaluate the correctness of an answer and award partial marks. The traditional use of multimedia content in education [5] emphasises subject knowledge but not cognitive skills. Testing students' broader range of skills is now possible thanks to the use of multimedia items. The traditional use of multimedia content in education emphasises subject knowledge but not cognitive skills.

### C. Analysis

The CROME aims for portability, resilience, scalability, and interoperability. The current implementation lays the groundwork for the development of a more robust, effective, and far-reaching multimedia architecture. Using the framework as a starting point, this paper aims to raise awareness among the multimedia research community and to encourage participation in topics that have received insufficient attention.

## V. NOSE SHAPE ESTIMATION AND TRACKING FOR MODEL BASED CODING

### A. Authors

Lijun Yin and Anup Basu

### B. Summary

Face feature extraction is critical in model-based coding [6] and human face recognition applications. Detecting and tracking the nose shape is not easy, and it is just as important as the eyes and mouth for model-based coding, particularly for the analysis and synthesis of realistic facial expressions. Experiments on real video sequences for low bit rate video coding show the benefit of the proposed scheme. Individual nostril and nose templates have been created. To reduce the sensitivity of facial feature detection to noise, a two-stage region growing algorithm - global region growing and local region growing - is developed, with a large growing threshold chosen to explore more skin area. The feature regions are extracted by segmenting the image using region growing on the colour components, namely luminance (Y) and chrominance (C) (U,V). Following the collection of feature blobs, they are classified using the k means classifier. To correctly detect the nostril shape, a geometric template, which is a twisted pair curve with a leaf-like shape, is applied to the nostril region.

### C. Analysis

The new method restricts the search region and employs the shape adaptive model to compensate for the shrinking effect of deformable templates. The deformable template matching algorithm fits the template to the facial features in a few single steps, avoiding the computational complexity of a large number of parameter searches and the template shrinking problem. The results demonstrate that the approach addressed in this paper is applicable in real-world applications.

## VI. VISUAL GESTURE RECOGNITION FOR GROUND AIR TRAFFIC CONTROL USING THE RADON TRANSFORM

### A. Authors

Meghna Singh, Mrinal Mandal, Anup Basu



Sample images showing (a) original data, (b) binary image data and (c) results of the medial skeletonization operation.

Fig. 3.

## B. Summary

The paper discusses a novel method for the recognition of hand gestures, used by air marshals for steering aircraft on the runway, using the Radon transform. Foreground is extracted from background by using an adaptive threshold value. Thinning is applied to images to obtain the medial skeleton. Fig. 3 shows the medial skeleton of each segment. [7] Radon Transform is used to detect orientation of skeletal lines. Once the Radon transform has been computed, we threshold the Radon transform coefficients to extract the most significant local regions. It is then normalized between 0-1. Finally, a k means clustering with  $k=8$  is applied for classification.

## C. Analysis

The proposed approach does not require any normalization of the image itself and promising recognition rates achieved with the parameterized Radon transform. However, it is observed that due to poor segmentation and thinning the medial skeleton is no longer a unitary segment but has loop like regions. These regions act as noise sources and cause a larger than usual Radon transformation. Thus when the Radon transform is thresholded these noise sources are not eliminated and get incorporated into the cumulative Radon transform and subsequently the normalized cumRT. These, noisy feature vectors sometimes lead to misclassification.

## VII. VARIABLE RESOLUTION TELECONFERENCING

### A. Authors

Anup Basu, Allan Sullivan and Kevin Wiebe

### B. Summary

This paper describes an image compression scheme based on Variable Resolution (VR) sensing. Compressed methods can be broadly classified as lossless [8] or lossy, depending on how the method affects the data being compressed [9]. The compression of motion pictures can be used in teleconferencing (video phones), electronic storage of movies, or even television transmission. The concept of variable resolution (VR) has been used in a variety of fields, but its application to image compression is novel. Images compressed with the VR transform can be compressed further using other

methods. A lookup table can be used to perform the VR transform. Any entropy technique's ability to compress data is dependent on the presence of patterns in the input data stream. Any entropy encoding technique's ability to compress data is dependent on the presence of patterns like repeating bytes or skewed frequency distributions, which digitized computer images lack. The VR transform has two parameters that influence the final image: the expected savings (compression) and alpha, which controls the distortion at the image's edges in relation to the fovea. VR cannot be used to compress all data because the distortions caused in many cases (such as commercial television broadcasts) would be unacceptable.

### C. Analysis

As an image compression technique [10], variable resolution has undeniable advantages. The rate at which it can compress images, particularly on slower machines, and the high quality present in the foveal region make it ideal for the teleconferencing market.

## VIII. STEREO MATCHING USING RANDOM WALKS

### A. Authors

Rui Shen, Iren Cheng, Xiabo Li and Anup Basu

### B. Summary

This paper introduces a new two phase stereo matching algorithm based on the random walks [11] framework, with promising results when tested on the Middlebury test bed. Stereo matching is a critical and traditional task in stereo vision, with the goal of determining the disparities of corresponding pixels in a pair of stereo images. A recent application of random walks in multi label image segmentation demonstrates great promise for its use in stereo matching. Prior matrices defined on the penalties of different disparity configurations and Laplacian matrices defined on pixel neighbourhood information are used to extract a set of reliable matching pixels. Following that, using the reliable set as seeds, the disparities of unreliable regions are calculated by solving a Dirichlet Problem. The algorithm operates in the (add that) space, and the parameters are automatically adjusted based on the lighting conditions of the input image pair. The main disadvantage highlighted in this paper is that only an exact and unique minimum solution of an energy function of the above form can be produced, whereas other approaches, such as Graph Cuts (GC), can only produce an approximation.

### C. Analysis

The algorithm was implemented in Mat Lab, and at least a tenfold speedup is expected once the author's algorithm is implemented in GPU, which will be significantly faster than many other global methods using a different optimization framework. The main advantage of random walks is that they can produce an exact and unique minimum solution of an energy function of the above, whereas other approaches, such as GC, can only produce an approximation. This paper not only proposed a new algorithm, but also demonstrated the

feasibility of a new framework for solving stereo matching, random walks [12]. The algorithm, which operates in the (add that formula) space, is limited to colour images.

## IX. EYE TRACKING AND ANIMATION FOR MPEG-4 CODING

### A. Authors

Stefan Bernogger, Lijun Yiny, Anup Basu, Axel Pinz

### B. Summary

The author of this research describes a method for synthesising eye movement by computing the deformation of the 3D model's eyes using the retrieved eye attributes. Create a 3D face model that is unique and map it to the first frame of the face sequence. The deformation parameters were used to the 3D model based on the retrieved eye attributes to synthesise the eye movement in successive frames. Determine two coarse zones of interest for the eyes in the eye tracking and detection algorithm. Using a gradient-based Hough transform, search the iris of the eyes. Select a small area of interest for obtaining the eye borders. Getting an initial approximation for the eye lids using colour information. Localize the eye lids using deformable templates. The eye detection and tracking algorithms extract the contours of the iris and eye lids in each frame of the image sequence. These eye features are used to synthesise the real motions of the eye on a 3D facial model.

### C. Analysis

The author found from experiments that the Hough transform [13], which is the first and therefore one of the most important steps, is very robust against noise as well as edges which are not produced by the contour of the iris. A high resolution facial model with 3118 patches is used to evaluate the algorithm. The proposed algorithms behaves well for synthesizing eye movements.

## X. PERCEPTUALLY OPTIMIZED 3D TRANSMISSION OVER WIRELESS NETWORKS

### A. Authors

Irene Cheng and Anup Basu

### B. Summary

The essential factors determining the display quality of three-dimensional objects (3D) are reviewed in this paper. The ability to adaptively adjust the model representation while maintaining satisfactory quality as perceived by a viewer is an important consideration in designing effective interactive online 3D systems. An overview of the research on developing a 3D perceptual quality metric that integrates two important ones, texture resolution and mesh resolution, which control the transmission bandwidth requirements. Simplification algorithms attempt to control mesh complexity by developing various strategies for simplifying the Level of Detail (LOD) in various parts of a 3-D object. One of the major drawbacks of most 3D transmission algorithms is that they do not account for data loss. Despite the fact that issues of multimedia

transmission over wireless networks have received attention, little work has been done on wireless 3D transmission [14]. Some preliminary implementations for deploying 3POPIT (3D Perceptually Optimized Partial Information Transmission) over a lossy wireless network [15] are highlighted in the paper. The two programs shown are combining and interpolating based on different texture and mesh sub samples, as well as a comparison of perpetually optimised versus non optimised transmission.

### C. Analysis

This paper discussed the factors that influence 3D image degradation and proposed a method for estimating perceptual quality while taking mesh and texture resolution variations into account. Following that, a strategy for optimising perceptual quality in the face of packet loss was presented. The approach is validated by the experimental results. Given partial packet transmission, the most computationally efficient approach to interpolation would be to predetermine neighbours and coefficients for interpolation and store various look up tables. Packet size and header length are not considered in the packet loss model, which is a simple model intended to demonstrate the feasibility of the method depicted.

## REFERENCES

- [1] C. D. Verestoy, J., "Tracking feature points: a new algorithm." In: *Proc. 14th Int. Conf. on Pattern Recognition*, pp. 1436–1438, Australia.
- [2] B. A. Murray, D., "Motion tracking with an active camera," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16 (5), pp. 449 – 459, 1994.
- [3] R. S. CHENG, I. and A. BASU, "Multimedia and games incorporating student modeling for education," In *Technology for Education, 2009. T4E '09. International Workshop on*, pp. 91 – 94, 2009.
- [4] M. Natarajan, "Use of online technology for multimedia education," *Information services & use*, vol. 26, no. 3, pp. 249–256, 2006.
- [5] B. B. Andresen and K. van den Brink, "Multimedia in education," in *Information technologies at school: conference materials*, 2002, pp. 5–8.
- [6] J. Ahlberg, "Model-based coding: Extraction, coding, and evaluation of face model parameters," Ph.D. dissertation, Linköpings universitet, 2002.
- [7] Y. A. N. Tsapatsoulis and S. Kollias, "On the use of radon transform for facial expression recognition," *Proc. of ISAS '99, Orlando, FL, USA*.
- [8] M. Yang and N. Bourbakis, "An overview of lossless digital image compression techniques," in *48th Midwest Symposium on Circuits and Systems, 2005*. IEEE, 2005, pp. 1099–1102.
- [9] M. Singh, S. Kumar, S. Singh, and M. Shrivastava, "Various image compression techniques: Lossy and lossless," *International Journal of Computer Applications*, vol. 142, no. 6, pp. 23–26, 2016.
- [10] M. Rabbani and P. W. Jones, *Digital image compression techniques*. SPIE press, 1991, vol. 7.
- [11] S. Karlin and J. McGregor, "Random walks," *Illinois Journal of Mathematics*, vol. 3, no. 1, pp. 66–81, 1959.
- [12] G. H. Weiss and R. J. Rubin, "Random walks: theory and selected applications," *Advances in Chemical Physics*, vol. 52, pp. 363–505, 1983.
- [13] G. Chow and X. Li, "Towards a system for automatic facial feature detection," *Pattern Recognition*, 26(12), p. 1739–1755, Dec 1993.
- [14] W. Lee, K. Noh, S. Kim, and J. Heo, "Efficient cooperative transmission for wireless 3d hd video transmission in 60ghz channel," *IEEE Transactions on Consumer Electronics*, vol. 56, no. 4, pp. 2481–2488, 2010.
- [15] M. Ghaderi, D. Towsley, and J. Kurose, "Reliability gain of network coding in lossy wireless networks," in *IEEE INFOCOM 2008-The 27th Conference on Computer Communications*. IEEE, 2008, pp. 2171–2179.