International Journal of Global Engineering (IJGE) E- ISSN: 2456-3099

VOL 2 ISSUE 1 (2017) PAGES 1 - 12

Received: 20/07/2017. Published: 18/08/2017

AUTOMATIC 2D-TO-3D IMAGE CONVERSION BASED ON DEPTH MAP ESTIMATION USING HAZE ALGORITHM AND PSEUDO DEPTH-MAP ESTIMATION

[1]Ms. A. Sivasankari, [2] Mrs. S. Ayisha Siddiqua, [3] Mrs. B. Arul Mozhi

[1]Head of the Department of Computer Science and Applications, DKM College for Women (A), Vellore, Tamil Nadu, India.

[2]Research Scholar, Department of Computer Science and Applications, DKM College for Women (A), Vellore, Tamil Nadu, India.

[3]Assistant Professor , Department of Computer science and Applications ,D.K.M college for Women(A) , Vellore, Tamil Nadu, India.

ABSTRACT

Recent advances in 3D have increased the importance of stereoscopic content creation and processing. Therefore, converting existing 2D contents into 3D contents is very important for growing 3D market. The most difficult task in 2D-to-3D conversion is estimating depth map from a single-view image. Thus, in this paper, we propose a novel algorithm to estimate the map by simulating haze as a global image feature. Besides, the visual artifacts of the synthesized left- and right-views can also be effectively eliminated by recovering the separation and loss of foreground objects in the proposed algorithm. Experimental results show that our algorithm can produce a good 3D stereoscopic effect and prevent the separation and loss artifacts with low computational complexity

Keywords: image, 2D-to-3D, conversion, depth estimation, separation, loss

1. INTRODUCTION

A rapid growth of commercialization of 3D display has increased the demands of 3D media contents for supporting full utilities of 3D displays and has aspired humans to experience more realistic and unique 3D effects. In addition to generating better visual experiences than conventional 2D displays, emerging 3D display have many applications, including movies, gaming, photograph, education and so on. However, due to lack of 3D media contents, converting existing 2D contents into 3D contents for growing 3D markets is very necessary and meaningful. How to generate or estimate the depth map using only a single-view image is the most important and difficult problem in 2D-to-3D conversion. Previous 2D-to-3D conversion methods are mainly divided into two classes: software-based method and depth cues-based method. The software-based method generates 3D content by using stereoscopic conversion tools, such as DDD's TriDef and ArcSoft's Media Converter, to retrieve depth maps. However, the stereoscopic visual effect produced by these tools is not obvious due to the limited information they used for conversion. A more feasible and effective method is the depth cues-based method. This kind of method is based on the key observation, that is, when observing the world, the human brain integrates various heuristic depth cues to generate the depth perception. The major depth perceptions are binocular depth cues from two eyes and monocular depth cues from a single eye [1]. The disparity of binocular visual system helps human eyes to converge and accommodate the object at the right distance. Monocular cues include focus/defocus, motion parallax, relative height/size, and texture gradient, providing various depth perceptions based on human experience. Therefore, humans can also perceive depth from the single-view image/video. The depth cues-based method assigns depth values using image classification [2], machine learning [3], depth from focus/defocus [4], depth from geometric perspective [5], depth from texture gradient, depth from relative height [6] and depth from multiscale local- and global-image features. For example, the computed image depth (CID) method [7] divides a single image into several sub-blocks and uses contrast and blurriness information to generate depth information for each block. Han, *et al.,* [8] generated the depth map by employing both vanishing points and super-pixels as geometric and texture cues. Cheng, *et al.,* [9] assigned the depth map based on a hypothesized depth gradient model. The method can produce impressive results. However, if the assumption of the global depth does not hold or large foreground objects exists, the method may fail in the cases. Yang, *et al.,* [10] generated a feasible perceptual depth map by using the local depth hypothesis that based on the structural information of the input image and salient regions. However, user interaction is required for this method. 2D-to-3D depth generation algorithms generally face two challenges. One is the depth uniformity inside the same object. The other challenge involves retrieving an appropriate depth relationship among all objects. Generating a depth map from single 2D images is an ill-posed problem. Not all the depth cues can be retrieved from an image. To overcome these two challenges, this work presents a novel algorithm that uses a haze veil to generate a pseudo depth map rather than retrieving the depth value directly from the depth cue. Firstly, the proposed algorithm produces a simulated haze image to represent salient region segmentation. Then the pseudo depth map is automatically generated in single-view image using the transmission information. Experimental results indicate that the proposed algorithm may generate promising stereoscopic results with slight side effects.An image veil is proposed in this paper to segment the saliency region from a single input image. This veil is generated based on the key observation that scene radiance is attenuated exponentially with depth, as indicated by the transmission map. If we can recover the transmission, then we can also recover depth information [11, 12]. Thus, depth information can be measured by the transmission map. To date, numerous studies have been done for estimating the transmission map from a haze image. Therefore, if we may transform a haze-free image into a haze image for the purpose of 2D to 3D conversion, then we may obtain depth information using various existing methods. A simple and effective method for removing haze is inspired by Retinex theory [13]. Based on this theory, the input image *I* with haze is the product of object reflectance *R* that can be regarded as haze-free image and scene illumination *L* that can be regarded as haze veil, that is:

$$I(x,y) = R(x,y) \cdot L(x,y) \quad . \qquad (1)$$

where (*x*, *y*) is the position coordinate of a pixel. The main idea of the haze removal algorithm is to estimate the haze veil with the mean of the illumination component *L* that is obtained by convoluting the haze input image with a zero mean Gaussian smoothing function *G*. This process can be written as follows

$$\hat{L}(x,y) = I(x,y) * G(x,y)\,, \tag{2}$$

$$\tilde{L}(x,y) = \frac{1}{\mathrm{HW}}\sum_{x=1}^{\mathrm{H}}\sum_{y=1}^{\mathrm{W}}\hat{L}(x,y)\,, \tag{3}$$

where L is the estimated haze veil, and H and W denote the height and weight of the image, respectively. The haze veil is subtracted from the original input image in the logarithmic domain to remove the haze effect from the input image, and then the exponential transformation is used to obtain the final haze-removed result R , as showed below

$$\tilde{r}(x,y) = \ln I(x,y) - \ln L(x,y) = i(x,y) - l(x,y)\,, \tag{4}$$

$$\tilde{R}(x,y) = \exp(\tilde{r}(x,y))\,. \tag{5}$$

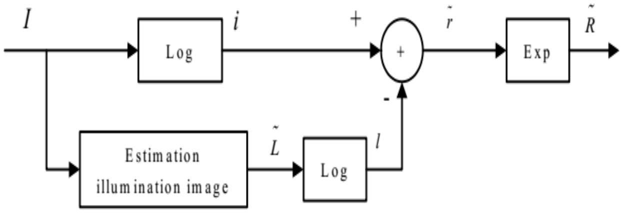


Figure 1. Flowchart of the Haze Removal Algorithm

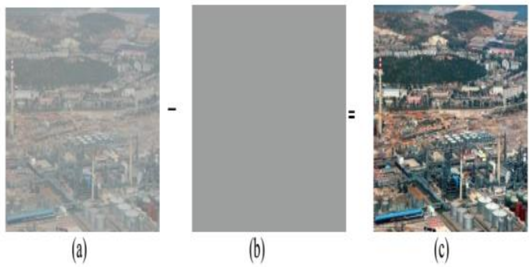


Figure 2. Illustration of the Haze Removal Procedure (a) Input Image (b) Estimated Haze Veil (c) Haze Removal Result

By creating a linear model for modeling the scene depth of the hazy image under this novel prior and learning the parameters of the model with a supervised learning method, the depth information can be well recovered. With the depth map of the hazy image, we can easily estimate the transmission and restore the scene radiance via the atmospheric scattering model, and thus effectively remove the haze from a single image and results show that the proposed approach outperforms state-of-the-art haze

Received: 20/07/2017. Published: 18/08/2017

removal algorithms in terms of both efficiency and the dehazing effect. Moreover, most automatic systems, which strongly depend on the definition of the input images, fail to work normally caused by the degraded images.

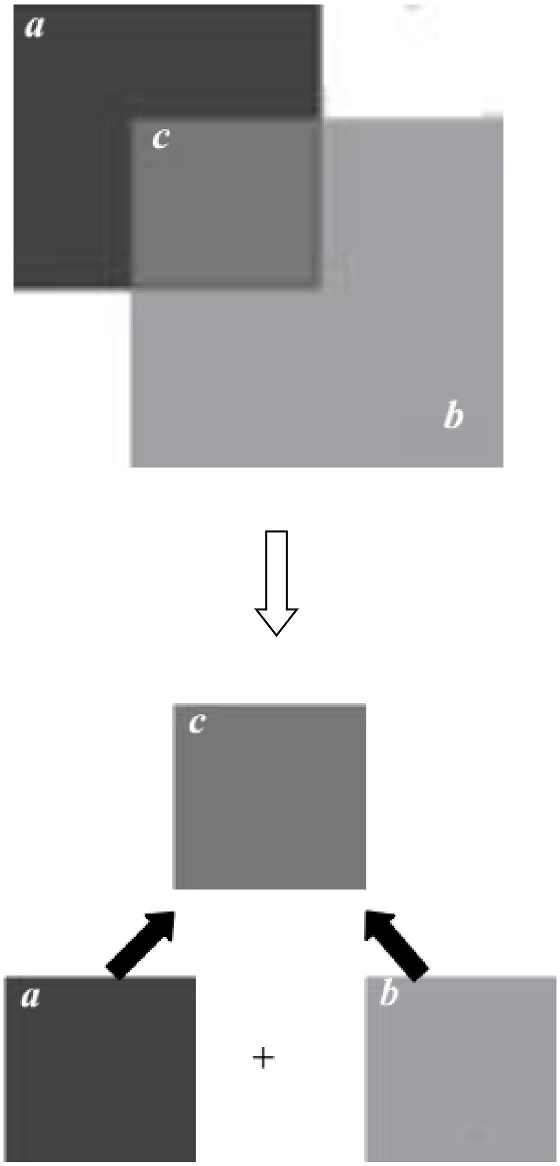


Figure 3 illustrates the haze illusion by using two blocks to create the haze effect. Thus, we can deduce that the haze image [see Figure 3(c)] is obtained by adding the haze-free input image [see Figure 3(a)] to the haze veil [see Figure 3(b)]. The haze effect is generated to compute the depth map of the input image.

2. PROPOSED ALGORITHM

2.1. Algorithm Procedure

Specifically, the proposed veil algorithm has three steps to automatically convert 2D image into 3D one. The first step is to generate a simulated haze image by adding a haze veil on the haze-free input image, and the haze image is used to represent salient region segmentation. The second step is computing the depth map by using the transmission map estimation in haze removal algorithm, which including initial depth map extraction, refined map estimation and final depth map estimation. The goal of the algorithm is to generate a depth map without using any heuristic depth cues or any user interaction. Finally, the 3D stereoscopic image is generated based on the estimated depth map. The overall procedure of this approach is depicted in Figure 4.

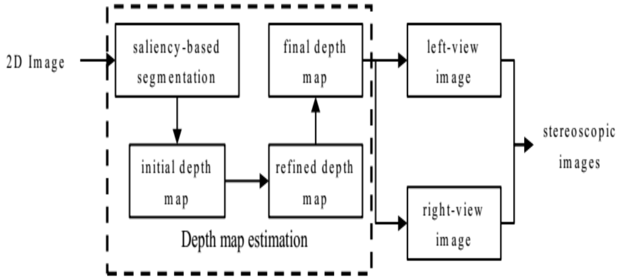


Figure 4. The Overall Procedure for 2D-To-3D Conversion

2.2. Pseudo Depth Map Estimation

In general, the 2D-to-3D conversion from a single image has been assigned to a problem of how to generate depth-map information from 2D images. The depth map estimation is automatic and consists in the following three stages: haze image simulation, initial depth map extraction, refined map estimation and final depth map estimation

2.2.1. Haze Image Simulation

In this section, we propose a method for simulating a haze image by adding a haze veil on the haze-free input image. The theory behind the haze simulation process is that if the haze veil can be subtracted from the degraded image to removal haze (see Figure 2), we can also simulate the haze image by adding the haze-free input image to the haze veil. In the haze removal experiments, we find that the veil estimated through a mean calculation of illumination component can only handle the uniform haze situation. If the haze is not uniform, the color distortion of haze removal result often occurs. However, it's not always true that the haze is evenly distributed at each position since the natural haze is dependent on the unknown depth information. Thus, we present a new way to estimate a non-uniform distributed haze veil in this paper.

According to the Koschmieder model [14], the apparent luminance of the scene objects at different distance is different, so different haze veil should be assigned according to their position. Therefore, we multiply the uniform veil *L* by the original image and apply the color inversion operation to obtain a depth-like map. Considering that the intensity of an image reflects the amounts of photons received by every position of an image, furthermore, the smaller the distance between the scene points and the camera, the stronger the intensity will be, thus the haze veil reflected by the depth-like map may be measured by its intensity. Therefore, we extract the intensity component of the depth-like map to produce the haze veil whose distribution is according to real fog density of the scene. Thus the haze veil for the input image, *L* ' is estimated by

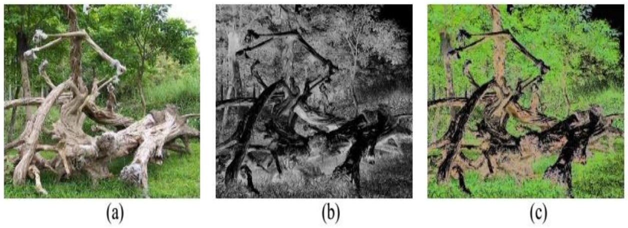
$$\tilde{L}'(x,y) = 255 - \omega_1 \times (R(x,y) \cdot \tilde{L}(x,y)) \,,$$

where *R* is the input image without haze, *L* is the mean of L obtained by Eq. (3) and □ $_1$ is an adjustment parameter set to 3 to generate a certain amount of haze in the input

image. Then, we transform the image *L* ' from RGB to YCbCr color space, and extract the intensity component of the image, which stands for our final haze veil. Once the depth-like haze veil *L* ' is figured out, the haze veil can be added on the real input haze-free image *R* to get the log-haze image *i* after the conditions are set. The process is expressed as follows

$$\tilde{i}(x,y) = \ln R(x,y) + \ln \tilde{L}'(x,y) \,.$$

Finally, the simulated haze image I_{haze} can be obtained using exponential transformation, that is I_{haze} □ exp (i (x , y)) . The saliency region is segmented from non-saliency regions (*e.g.*, the sky and objects or surfaces that are too dark or too light) in the image I_{haze}, such that haze image simulation is actually the image segmentation based on saliency. For example, Figure 5(a) and Figure 5(b) show the original 2D image and the estimated haze veil. The simulated haze image is shown in Figure 5(c).


(a) (b) (c)

Received: 20/07/2017. Published: 18/08/2017

Figure 5. Process of Simulating Haze Image (a) Original 2D Image (b) Haze Veil (c) Simulated Haze Image

2.2.2. Pseudo Depth-map Estimation

Once the haze image I_{haze} is obtained, we can adopt the transmission estimation method that widely used in haze removal to obtain depth information. For this purpose, the dark channel prior [11, 12] and a guided filter [15] are used to estimate the depth map.

Specifically, we first estimate the atmospheric light A for the image I_{haze}. Most algorithms estimate A from the pixels with highest intensities, which is fast but not accurate. He, *et al.*, [11, 12] integrate the atmospheric light estimation with the dark channel prior and it makes the estimation result more accurate. This method is also adopted in this paper.

The depth map is calculated based on the image degradation model [14] and the dark channel prior proposed by He [11, 12]. For the haze image, we first estimate the initial depth map $m\ (\ x\ ,\ y\ )$, this process can be written as

$$\tilde{m}(x,y) = 1 - \omega_2 \min_{c\in\{R,G,B\}}\left(\min_{(x',y')\in\Omega(x,y)}\left(\frac{I_{haze}^{c}(x',y')}{A^{c}}\right)\right)$$

where $I\ {}_{h}^{c}\ {}_{a\ z\ e}$ is a color channel of I_{haze} , □ (x , y) is a local patch centered at (x, y), and (x ', y ') is the pixel location that belong to □ (x , y) . □ $_2$ is a constant parameter for adjusting the amount of haze for distant objects. The value of □ $_2$ is set to be 0.95 for all the results
reported in this paper.

It should be noticed that there are obvious block effects and redundant details in the initial depth map. In order to handle these deficiencies, we thus use the guided filter [15] and bilateral filter to refine the initial depth map. The detailed estimation process of the final depth-map is described in the following steps.

Step 1. For the initial depth map, we first compute the linear coefficients a_k and b_k for the guided filter

$$a_k = \frac{\frac{1}{|\omega|}\sum_{(x,y)\in\omega_k} I_{haze}(x,y)\tilde{m}(x,y) - u_k\bar{m}_k}{\sigma_k^2 + \varepsilon} \qquad (9)$$

$$b_k = \bar{m}_k - a_k u_k$$

where I_{haze} is the guidance image and m is the input image of the guided filter since the filter is a general linear translation-variant filtering process, which involves a guidance image and an input image [11]. In Eq. (9), □ is a regularization parameter preserving a_k from being too large. u_k are the mean and variance of I_{haze} in a window that

Step 2. Once the linear coefficients (a_k, b_k) are obtained, we can compute the filter output by

$$m'(x,y)=\bar{a}_k\tilde{m}(x,y)+\bar{b}_k \qquad (10)$$

Step 3. A bilateral filter is used here to remove the redundant details for the refined depth map m' since the bilateral filter can smooth images while preserving edges. Thus, the redundant details of the refined depth map m' estimated by the algorithm presented above can be effectively removed. This process can be written as

$$\hat{m}(\mathbf{u})=\frac{\sum_{\mathbf{p}\in N(\mathbf{u})} W_c(\|\mathbf{p}-\mathbf{u}\|)W_s(|m'(\mathbf{u})-m'(\mathbf{p})|)m'(\mathbf{p})}{\sum_{\mathbf{p}\in N(\mathbf{u})} W_c(\|\mathbf{p}-\mathbf{u}\|)W_s(|m'(\mathbf{u})-m'(\mathbf{p})|)}$$

where $m'(\mathbf{u})$ is the refined depth map corresponding to the pixel $\mathbf{u}=(x, y)$, $N(\mathbf{u})$ is the neighbors of $\mathbf{u}$. The spatial domain similarity function $W_c(x)$ is a Gaussian filter with the standard deviation is □ $_c$: $W_c(x)$ □ e $^{□ x 2 / 2 □ c2}$, and the intensity similarity function $W_s(x)$ is a Gaussian filter with the standard deviation is □ $_s$, it can be defined as: $W_s(x)$ □ e $^{□ x 2 / 2 □ s2}$. In our experiments, the value of □ $_c$ and□ $_s$ is set as 3 and 0.4, respectively. Thus, we can obtain the final depth map $\hat{m}(x, y)$.

Figure 6 shows the corresponding initial depth map, refined depth map and the final depth map for the original image in Figure 5(a). From these figures, one can see that the final depth map [see Figure 6(c)] generated using the proposed method reflects the relative positions between scene objects and their neighboring regions. Thus, the map is a pseudo depth map instead of a recovery of real depth information. Generally, the pseudo map is based on the visual attention of mapping the saliency regions from the position close to the viewer while mapping the non-saliency regions from farther positions.

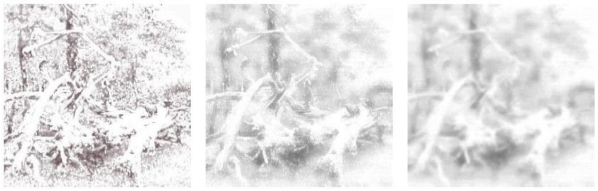


Figure 6. Process of Estimating the Depth Map (a) Initial Depth Map (b) Refined Depth Map (c) Final Depth Map

2.3 3D Image Visualization using Depth Map-based Rendering

Received: 20/07/2017. Published: 18/08/2017

Once the depth map is obtained, the left-view and the right-view images can be synthesized by the following steps. Firstly, we compute the parallax value *Parallax*(*x*, *y*) from each pixel (*x*, *y*) in the estimated depth map. The computation of the parallax value can be written as

$$Parallax(x,y) = \omega_3 \times \left(1 - \frac{\hat{m}(x,y)}{\mathrm{ZPP}}\right),$$

where *m* (*x* , *y* is the final depth map for the single image, □ $_3$ is the maximum parallax value. As can be seen in Figure 7(a), we can get principle.ocular distance *E* is about 6.35cm. *D* is the Max depth into the screen, and it is set to 10cm. Thus, the computed □ $_3$ value is 0.578cm. Next, we should express the value □ $_3$ in the form of pixel. In our experiment, 17’’ monitor (1280 □ 1024 Resolutions) is used here, so 1cm on the monitor is corresponding to 38 pixels. Thus, the maximum parallax value □ $_3$ is approximately 30 pixels for the image having a width size approximate to 1000. The zero parallax plane (ZPP) is set as the region with the depth value of *Th*, which is computed by *T h* □ m a x (*m*ˆ (*x* , *y*)) □ 1 0 to prevent separation and loss of artifacts.

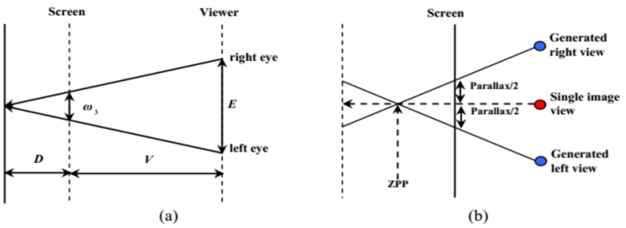
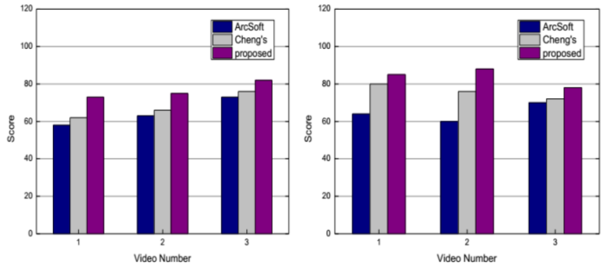


Figure 7. Stereoscopic Generation (a) Max Parallax Computation (b) Right View and Left View Generation



(a) Depth Quality (b) Visual Comfort

Figure 7. Quantitative Evaluation Results for the Test Videos

A novel and automatic method was proposed to generate a pseudo depth map in single view image using the estimated haze veil. A haze image was simulated by adding a haze veil on the input image to represent salient region segmentation, and then it estimated

Received: 20/07/2017. Published: 18/08/2017

pseudo depth map by using the transmission estimation method in haze removal algorithm.

CONCLUSION

Thus using the depth map, left- and right-view images were synthesized, and finally the stereoscopic images were generated to provide a sense of depth to the viewers with the help of anaglyph glasses. One can clearly see that the proposed algorithm has the best scores in depth quality and visual comfort. This confirms our observations , so all the human can have depth perception with daily life experience. Besides, the separation and loss artifacts of the synthesized results can also be effectively prevented with low computational complexity. The whole process of the proposed algorithm without any heuristic cues or user interaction. This work includes improving the computational efficiency of the proposed algorithm using GPU implementation or parallel computation.

References

[1] W. J. Tam and L. Zhang, "3D-TV content generation: 2D-to-3D conversion", Proceedings of IEEE International Conference on Multimedia and Expo, **(2006)** July 9-12, Toronto, Canada.

[2] S. Battiato, S. Curti, E. Scordato, M. Tortora and M. La Cascia, "Three-Dimentional Image Capture and Applications V1", vol. 5302, **(2004)**.

[3] D. Hoiem, A. A. Efros and M. Hebert, ACM Transactions on Graphics, vol. 3, no. 24, **(2005)**.

[4] J. Park and C. Kim, Proceedings of SPIE, vol. 6077, **(2006)**.

[5] Y.-M. Tsai, Y.-L. Chang and L.-G. Chen, "Block-based vanishing line and vanishing point detection for 3d scene reconstruction", Proceedings of International Symposium on Intelligent Signal Processing and
Communications, **(2006)** December 12-15, Yonago, Japan
.

[6] Y. J. Jung, A. Baik, J. Kim and D. Park, "A novel 2D-to-3D conversion technique based on relative height depth cue", Proceedings of SPIE - The International Society for Optical Engineering, **(2009)** January 19-21, San Jose, United states.

[7] H. Murata, SID Digest of Technical Papers, vol. 1, no. 29, **(2014)**

[8] Digital Imaging and Communications in Medicine (DICOM) Part 5: Data Structures and Encoding Published by National Electrical Manufacturers Association 1300 N. 17th Street Rosslyn, Virginia 2009 USA.

Received: 20/07/2017. Published: 18/08/2017

[9] F. Tsalakanidou and S. Malassiotis, "Real-time 2D+3D facial action and expression recognition,"Pattern Recognition, vol. 43, no. 5, pp. 1763–1775, 2010.

[10] M. C. Fairhurst, "New perspectives in automatic signature verification," Information Security Technical Report, vol. 3, no. 1, pp. 52–59, 2008.

[11] K. Bashir, T. Xiang, and S. Gong, "Gait recognition without subject cooperation," Pattern Recognition Letters, vol. 31, no. 13, pp. 2052–2060, 2010

[12] C. M. Verwoerd-Dikkeboom, A. H. J. Koning, W. C. Hop, P. J. van der Spek, N. Exalto, and E. A. P.Steegers, "Innovative virtual reality measurements for embryonic growth and development,"Human Reproduction, vol. 25, no. 6, pp. 1404–1410, 2010.

[13] Kaiming H, Sun J, Tang X (2011) Single image haze removal using dark channel prior. Pattern Analysis and Machine Intelligence, IEEE Transactions 12: 2341-2353

[14] Zichong C, Feng Y, Lindner A, Barrenetxea G, Vetterli M (2012) How is the weather: Automatic inference from images. 19th IEEE International Conference on Image Processing (ICIP) 1853-1856.

[15] Tarel JP, Hautière N, Gruyer D, Cord A, Halmaoui H (2010) Improved visibility of road scene images under heterogeneous fog. Intelligent Vehicles Symposium (IV), IEEE 478-485

[16] Xunshi Y, Luo Y, Zheng X (2009) Weather recognition based on images captured by vision system in vehicle. Advances in Neural Networks, CISNN, Springer Berlin, Heidelberg 390-398.

[17] M.A. Fischler and R.C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Graphics and Image Processing, 24(6):381 – 395, June 2007.

[18] R.I. Hartley. Euclidean reconstruction from uncalibrated views. In Proceeding of the DARPA–ESPRIT workshop on Applications of Invariants in Computer Vision, Azores, Portugal, pages 187–202, October 2003.

[19] Ricardo L. de Queiroz, Senior Member, IEEE, and Karen M. Braun, ―Color to Gray and Back: Color Embedding into Textured Gray Images‖, IEEE transactions on image processing, Volume.15, No. 6, June2010.

[20] J. Gomes and O. Faugeras. Reconciling Distance Functions and Level Sets. Journal of Visual Communication and Image Representation, 11:209-223, 2011.

[21] Y. Bai, S. J. Harrington, and J. Taber, "Improved algorithmic mapping of color to texture," Proc.

SPIE, vol. 4300, pp. 444–451, 2010.

[22] R. L. de Queiroz, Compression of Color Images, in the Handbook onTransforms and Data Compression, G. Sharma, CRC 2012.

Ms. A. SIVASANKARI M.Sc.,M.Phil.,DCP.,
Head of the Department
Department of Computer science and Applications,
D.K.M College for women (Autonomous),
Vellore Tamil Nadu India.

Mrs. S.AYISHA SIDDIQUA
Research Scholar
Department of Computer science and Applications,
D.K.M College for women (Autonomous),
Vellore Tamil Nadu India.

Mrs. B. ARUL MOZHI M.Sc.,B.Ed.,
M.Phil., SET.,
Assistant Professor,
Department of Computer Science and
Applications D.K.M College women(A),
Vellore ,Tamil Nadu, India.

.