

Reconstruction of 3d video from 2d real-life sequences

Reconstrucción de video 3d desde secuencias reales en 2d

*Eduardo Ramos Diaz**, *Volodymyr Ponomaryov*

National Polytechnic Institute of Mexico, E.S.I.M.E.-Culhuacan, Av. Santa Ana N.º 1000, Col. San Francisco Culhuacán, C.P.04430, Mexico City, Mexico

(Recibido el 27 de junio de 2009. Aceptado el 18 de mayo de 2010)

Abstract

In this paper, a novel method that permits to generate 3D video sequences using 2D real-life sequences is proposed. Reconstruction of 3D video sequence is realized using depth map computation and anaglyph synthesis. The depth map is formed employing the stereo matching technique based on global error energy minimization with smoothing functions. The anaglyph construction is implemented using the red component alignment interpolating the previously formed depth map. Additionally, the depth map transformation is realized in order to reduce the dynamic range of the disparity values, minimizing ghosting and enhancing color preservation. Several real-life color video sequences that contain different types of motions, such as translational, rotational, zoom and combination of previous ones are used demonstrating good visual performance of the proposed 3D video sequence reconstruction.

----- *Keywords:* Video sequence, anaglyph, depth map, dynamic range

Resumen

En este artículo, se propone un método novedoso que permite generar secuencias de video en 3D usando secuencias de video reales en 2D. La reconstrucción de la secuencia de video en 3D se realiza usando el cálculo del mapa de profundidad y la síntesis de anaglifos. El mapa de profundidad es formado usando la técnica de correspondencia estéreo basada en la minimización de la energía de error global a partir de funciones de suavizado. La construcción del anaglifo es realizada usando la alineación del componente de color interpolándolo con el mapa de profundidad previamente formado. Adicionalmente, se emplea la transformación del mapa de profundidad para

* Autor de correspondencia: teléfono: 52 + 55 + 565 620 58, fax: + 52 + 55 + 565 620 58, correo electrónico: eramos@ieee.org. (E. Ramos)

reducir el rango dinámico de los valores de disparidad, minimizando el efecto fantasma mejorando la preservación de color. Se usaron numerosas secuencias de video a color reales que contienen diferentes tipos de movimientos como traslacional, rotacional, acercamiento, y la combinación de los anteriores, demostrando buen funcionamiento visual de la reconstrucción de secuencias de video en 3D propuesta.

----- *Palabras clave:* Secuencia de video, Anaglifo, Mapa de profundidad, Rango dinámico

Introduction

In recent years, we are seeing great advances in the development of different techniques to reconstruct 3D video sequences from 2D ones. Significant advances have also occurred in the development of stereoscopic displays. Such efforts enable the 3D viewing and multi-view autostereoscopic displays that show more than two view points for a given scene [1, 2].

While the advances above mentioned are more evident, the problem of content generation still lingers. Additionally, stereoscopic displays are still expensive in comparison to conventional displays.

One possible solution to this problem is conversion of 2D to 3D video using optical flow, depth map, or other techniques to extract image information that should be used later to reconstruct the 3D video sequence.

Anaglyph is considered the simplest way to obtain 3D perception and have less computational cost in comparison with other methods, such as Photoshop, and the least squares algorithm [3].

In order to construct anaglyphs, the simplest method relies on simple time delay between frames and adjustment of left-right images [4].

Usually, to obtain anaglyphs, the left view in blue (or green) color is superimposed on the same image with the right view in red one. When viewed through spectacles of corresponding colors but reversed, the three-dimensional effect can be perceived.

Anaglyph based 3D projection is considered as the cheapest stereoscopic projection technique, because can be implemented at television sets or computer screens with no special hardware different than inexpensive colored glasses. In this case, it can be used a color anaglyph based on the Photoshop for red-cyan glasses presented, for example as in [3].

Recently proposed in [3] approach can be applied that employs the adjacent frames from video sequence in order to compute depth maps. The resulting depth map is used to generate synthetic views. These depth maps can be also employed to reconstruct 3D images, because they contain the information on the 3D shape of the scene.

Other simple algorithm to construct anaglyphs is presented in [5], where the anaglyph is realized in the depth map acquisition from MPEG-4 protocol.

There are several methods to compute depth maps from a stereo pair or via adjustment of the video frames, among them, the optical flow algorithms, matching algorithms, etc. The major drawback of all these methods is that they require intensive computational operations.

In papers [6, 7], the differential techniques Lucas and Kanade, and Horn and Shunck to compute optical flow are presented, where it has been implemented a weighted least squares fit of local first order constraints to a constant model for motion map in each small spatial neighborhood.

In papers [8, 9], the energy-based method on the output energy of velocity-tuned filters is

applied, where the local energy is extracted using twelve Gabor-energy filters of several spatial scales, tuned to different spatial orientations and different temporal frequencies. Calculating a dense depth map is basically consisted of finding the correspondence between a pixel location in one frame and its position in the other one. This mapping of an image to other one can be obtained by registering a spatial neighborhood, surrounding each a pixel in an image to another one.

Stereovision technique is one of the methods yielding depth information of the scene. It uses the stereo image pairs from two cameras or a video sequence [10] to produce disparity maps that can be easily turned into depth maps.

Most disparity estimation algorithms that exist are pixel-based ones [10-12], in which the disparities are estimated pixel-by-pixel, and that is one a common feature for various otherwise quite different disparity estimation algorithms. The main problem associated with pixel-based approaches is that they cannot effectively handle un-textured surfaces.

In work [13], the mean shift segmentation algorithm is used to segment the images into different regions, then, the depth map is calculated.

Region based disparity estimation uses the stereo image pairs from two cameras to produce disparity maps that can be easily turned into dept maps. Reliability of depth maps or optical flow computation and computational cost of algorithm is key issue for implementing robust applications.

In this work, a novel algorithm to reconstruct 3D video sequences from 2D information is presented.

The paper is organized as follows. The depth map computation, the anaglyph generation and improvement, and data used are presented in section 1. In section 2, we expose the simulation results for depth map, anaglyph synthesis, anaglyph enhancement, and a brief discussion of the results. Finally, the conclusion of this work is presented in section 3.

Methodology

In order to obtain the 3D video, the following steps are should be employed: depth map computation, anaglyph construction, anaglyph enhancement and finally, the 3D video construction from anaglyph obtained. Below each a step is described in details.

Depth map computation

Using the classic definition of stereo projection, the depth can be calculated as follows:

$$Z(i, j) = f \cdot \frac{T}{d(i, j)} \quad (1)$$

Where T is the distance between camera lens, f is the focal length, and $d(i, j)$ is the disparity map between two images.

We use the method proposed in paper [14], which is based on the Global Error Energy Minimization by Smoothing Functions.

To calculate error energy matrix for every disparity value, the block matching technique has been implemented using different window sizes.

Let denote left image as $L(i, j, c)$ and right one as $R(i, j, c)$, each one in RGB format, so, the error energy can be written as:

$$e(i, j, d) = \frac{1}{3 \cdot n \cdot m} \sum_{x=i}^{i+n} \sum_{y=j}^{j+m} \sum_{k=1}^3 (L(x, y + d, k) - R(x, y, k))^2 \quad (2)$$

Where d is the disparity, c represents color components in an image, m and n are the rows and columns of an image.

For a predetermined range of disparity search range, an averaging filter

$$\tilde{e}(i, j, d) = \frac{1}{n \cdot m} \sum_{x=i}^{i+n} \sum_{y=j}^{j+m} e(x, y, d) \quad (3)$$

is applied several times to smooth the error in the energy matrix of disparity removing very sharp

changes in energy, which possibly belongs to incorrect matching.

We select the disparity that has the minimum error energy as the most reliable disparity estimation of the disparity map. So, the error energy of disparity maps can be written as:

$$E_d(i, j) = \frac{1}{3 \cdot n \cdot m} \sum_{x=i}^{i+n} \sum_{y=j}^{j+m} \sum_{k=1}^3 (L(x, y + d(i, j), k) - R(x, y, k))^2 \quad (4)$$

and the reliability of the obtained disparity map is calculated as follows:

$$R_d = \frac{1}{S_d} \left(\sum_{(i,j)-E_d(i,j)=ne}^{n,m} E_d(i, j) \right) \quad (5)$$

where S_d is the number of points in error energy, which are no-estimated (ne).

Disparity map contains some unreliable disparity estimation for some points, so, to avoid this, a threshold should be applied in such a form:

$$\tilde{d}(i, j) = \begin{cases} d(i, j), & E_d(i, j) \leq V_e \\ 0, & E_d(i, j) > V_e \end{cases} \quad (6)$$

$$\tilde{E}_d(i, j) = \begin{cases} E_d(i, j), & E_d(i, j) \leq V_e \\ ne & E_d(i, j) > V_e \end{cases}$$

Here, we use V_e as an error energy threshold, that limits the disparity map, and is found as: $V_e = \alpha \cdot Mean(E_d)$, where α is the tolerance coefficient to adjust the filtering process reliability, and $0 < \alpha < 1$.

Anaglyph generation and enhancement

Once the disparity map is computed [15], we can construct the depth map as follows:

$$Depth Map(i, j) = c \sqrt{MV_x^2(i, j) + MV_y^2(i, j)} \quad (7)$$

where c is used as a scaled parameter to fits the maximal disparity over the frames; $MV(i, j)_x^2$ and $MV(i, j)_y^2$ are the X and Y motion vectors values for each a pixel.

Classical methods, such as Photoshop, least squares, etc. which are used in anaglyph construction, always produce ghosting effects and color losses. Therefore, dynamic range reduction of the depth map values could be employed to minimize mentioned drawbacks. Using the P-th law transformation for dynamic range compression [16], the original depth map D is changed as follows:

$$D_{new} = a \cdot D^P \quad (8)$$

where D_{new} is the new depth map pixel value, $0 < a < 1$ is a normalizing constant, and $0 < P < 1$. The dynamic range compression permits retaining the depth ordering information within the depth map, while reducing ghost effects in the non-overlapping areas in the anaglyph.

Anaglyph synthesis requires generating a new stereo pair from one of the images and the modified map, so, the initial stereo pair image has to be re-sampled in a grid dictated by the depth map. This process should be realized via an interpolation. Some interpolation methods, such as bilinear, sp-line and nearest neighbor are tested experimentally, but in exposed experiments we only use the nearest neighbor interpolation as a promising one according to results obtained.

The nearest neighbor interpolation uses the neighborhoods defined by pixels (u, v) , $(u, v + 1)$, $(u + 1, v)$, $(u + 1, v + 1)$ around the point (X, Y) . Among the four pixels, the one closest to (X, Y) is determined, and its intensity is used as intensity for (X, Y) point. A flow chart of design the 3D video sequence proposed here is exposed in figure 1. Firstly, the video sequence decomposition, in order to obtain frames, is arranged into pairs; secondly, each a color component should be separated. Then, it is necessary to compute depth map and realize the dynamic range manipulation in order to reduce

ghosting effects; therefore, the nearest neighbor interpolation for the compressed depth map and red component in each a pair of the frames is required. The interpolation method provides the color preservation and low computational cost in the video processing. Finally, the reconstructed 3D video is formed. We apply all process presented below to each an image pair through a video sequence.

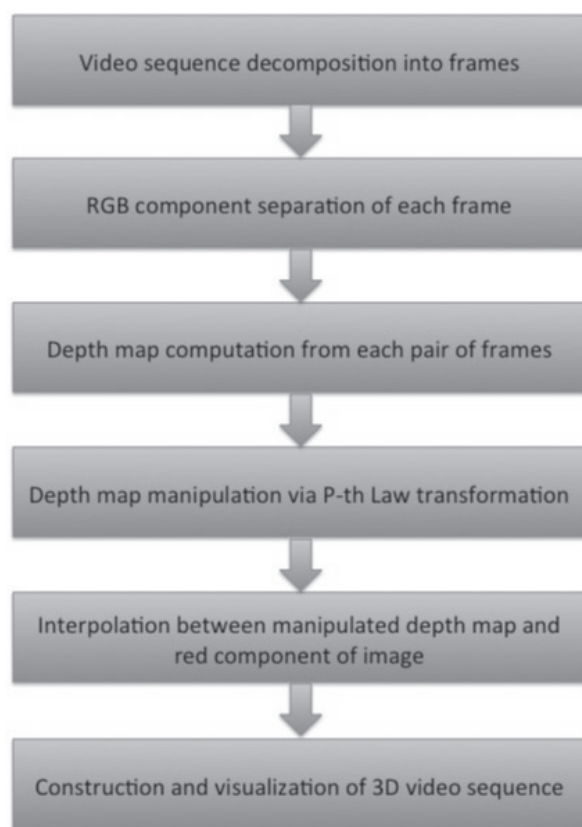


Figure 1 Flow chart to obtain 3D video sequence

Data used

In the simulation experiments, different real-life color video sequences, such as *Coastguard*, *Flowers*, *Foreman*, *Salesman* and *Alley* are employed in the Avi format. *Coastguard* color video sequence (300 frames, frame size 352 x

288 pixels) presents a particular boat movement from left to right side in the image; the *Flowers* sequence (300 frames, 352 x 288 pixels) exposes the right to left camera movement; in *Foreman* sequence (300 frames, 352 x 288 pixels), one can see a talking man; the *Salesman* sequence (48 frames, 352 x 288 pixels) shows the man that, holding a little box in his hands, realizes some movements; and finally, the *Alley* sequence (50 frames, 336 x 272 pixels) presents a movement as if we are walking in the alley. In figure 2, we present the images extracted from video sequences that used as stereo pair further.

Results and discussion

In this section, the obtained experimental results are presented: the depth map calculation using the proposed method; anaglyph results, which are obtained for the constructed anaglyphs with nearest neighbor interpolation method; and finally, the depth map manipulated via dynamic range is used to the anaglyph reconstruction. Simulation data were acquired varying the parameters in the depth map calculation in order to realize the better perception in 3D. The video sequence was separated in JPEG images, allowing the usage of the frames as stereo pairs.

Depth map

The video sequence is separated into frames and disparity maps are computed for each a frame using the corresponding stereo pair. Then, the corresponding depth map for each a frame from all video sequence is computed using MatlabTM R2009a software with an AMDTM Phenom X3 processor (64 bits) in a PC.

Figure 3 exposes the depth map image, where white pixels represent pixels with movement and black pixels represent pixels without movement into a frame. In all the cases, the obtained depth map of the stereo pair is sufficient for 3D perception purpose.

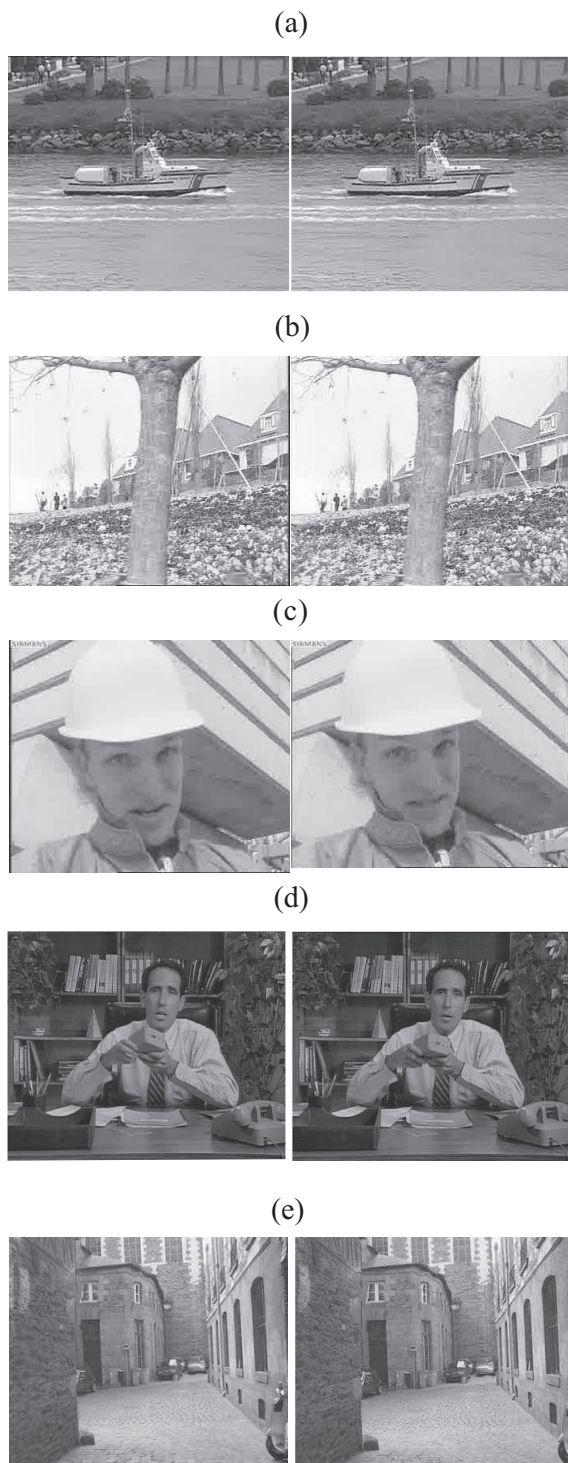


Figure 2 Video frames (neighboring frames used as the stereo pairs) for the following sequences: a) *Coastguard*, b) *Flowers*, c) *Foreman*, d) *Salesman* and e) *Alley*

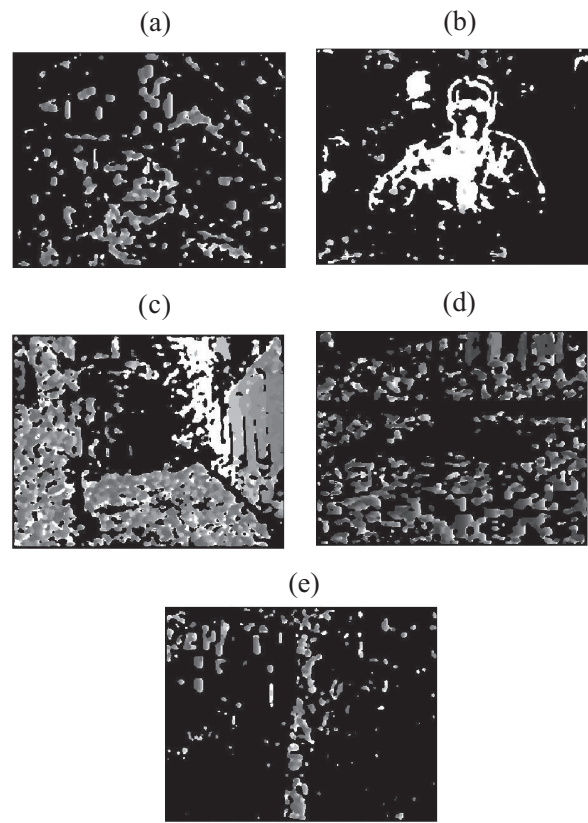


Figure 3 Resulting depth map for the following sequences: a) *Coastguard*, b) *Flowers*, c) *Foreman*, d) *Salesman* and e) *Alley*

Anaglyph synthesis

Once depth map is computed, anaglyph synthesis can be realized by the interpolation of the red component of the right image using the depth map. At the interpolation stage, nearest neighbor method has been employed because it is capable to produce satisfactory results in the visual observation.

Figure 4 presents the anaglyph synthesized for each a video sequence. In left images, the anaglyph construction is exposed using nearest neighbor Interpolation that does not significantly reduce the quality of the resulting 3D video in comparison with other interpolation algorithms [17, 18]. In the right images, the resulting anaglyph with depth map compression is shown, when the P-th law transformation is applied to depth map in order to compress the disparity values [19]. Values

P=0.5 and a=0.5 were selected for all the frames in the investigated video sequences. As one can see, the application of the compression improves

the anaglyph and allows better 3D perception, reducing ghosting in the final anaglyph. The selected parameter values are shown in table 1.

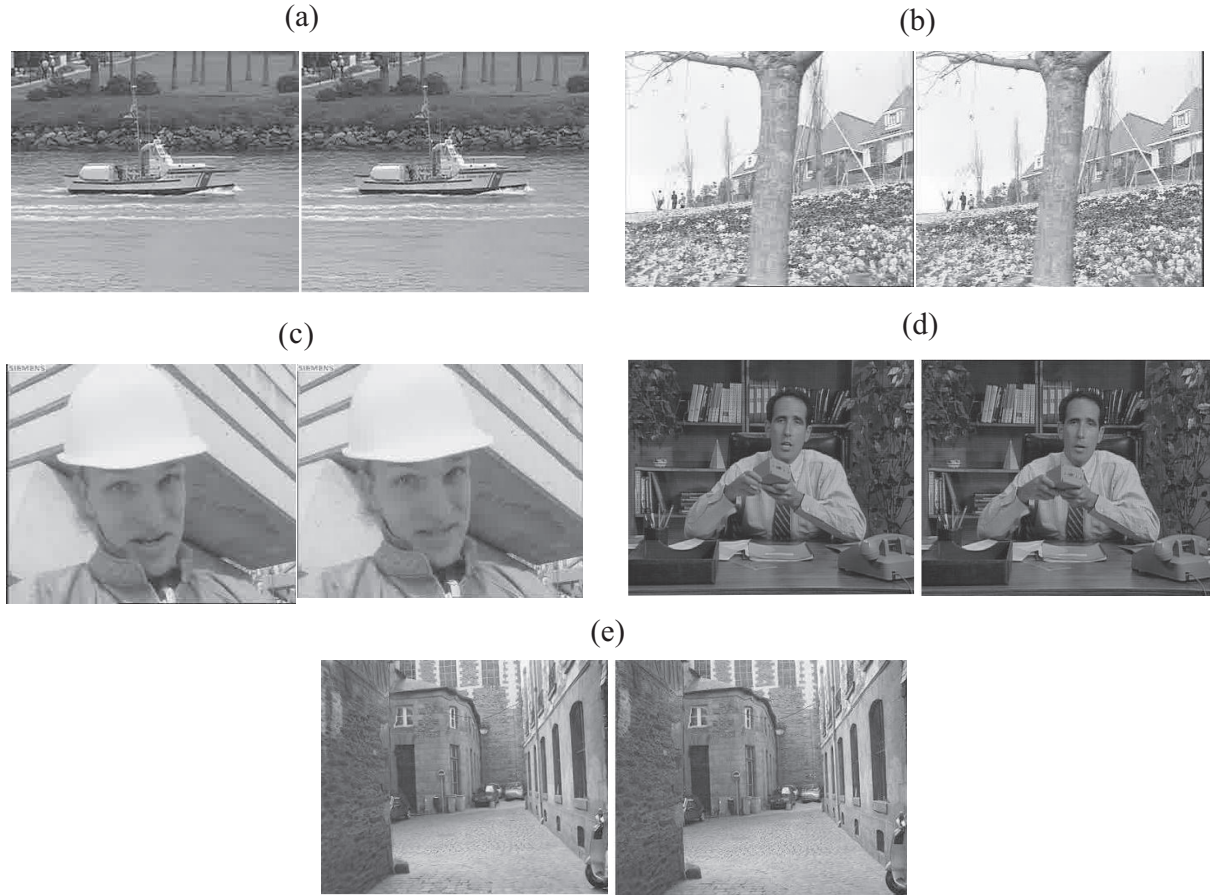


Figure 4 Synthesized Anaglyph without depth map manipulated (left images), and synthesized anaglyph with depth map adjusted using P-th law compression (right images) for: a) *Coastguard*, b) *Flowers*, c) *Foreman*, d) *Salesman* and e) *Alley*.

Table 1 Selected parameter values for each a video sequence and calculating time

Sequence	Alpha	Disparity threshold	Disparity map	Disparity map filtered	Focal lenght	Calculating time (sec)
Coastguard	1	2.2	0.0057	0.0022	40	26.09
Flowers	0.9	2.0	0.0036	0.00081	40	21.89
Foreman	0.9	2.0	0.030	0.0042	40	14.80
Salesman	1	5.0	0.051	0.013	40	14.72
Alley	1	0.6	0.003	0.0011	60	13.11

Calculating time is obtained since the video capture even the anaglyph visualization in a standard monitor.

In order to compare the proposed technique with a commonly used one in reconstruction of 3D video sequences, we implement the Lucas & Kanade classic differential technique to compute optical flow [8], constructing the depth map according to equation (7).

The figure 5 presents the depth map results using Lucas and Kanade technique justifying

the possibility to construct the similar depth map employing the Lucas and Kanade method, where one can see its drawback, missing of some details, which lead to worse results in 3D view using anaglyphs. Depth map computation via region based stereo matching calculation contains more important motion elements that are not presented in Lucas and Kanade algorithm, so, this justifies that the proposed algorithm allows better 3D perception without significant loss information extracted from stereo pair.

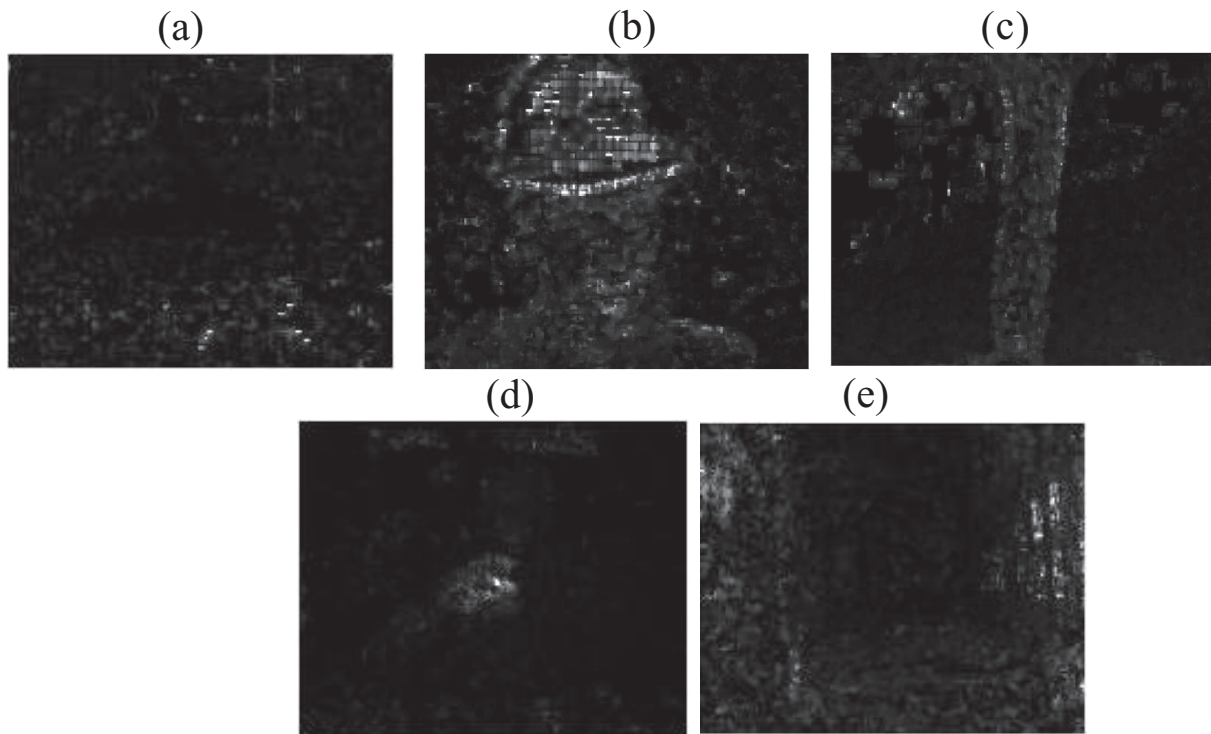


Figure 5 Depth Map computation using Lucas & Kanade technique for: a) *Coastguard*, b) *Flowers*, c) *Foreman*, d) *Salesman* and e) *Alley*

Additionally, in order to verify that the 3D perception is worth in the case of the Lucas and Kanade depth map computation, we used it to synthesize the corresponding anaglyph. The figure 6 shows the obtained results.

Observing the figure 6 with spectacles, the ghosting effects and low 3D perception are easy observed in

That can be explained as a consequence of poor details' reconstruction in the depth map formed by optical flow technique.

In order to compare the proposed method to the anaglyph reconstruction, we implemented the original Photoshop (PS) algorithm presented in [20] exposing the results in figure 7.

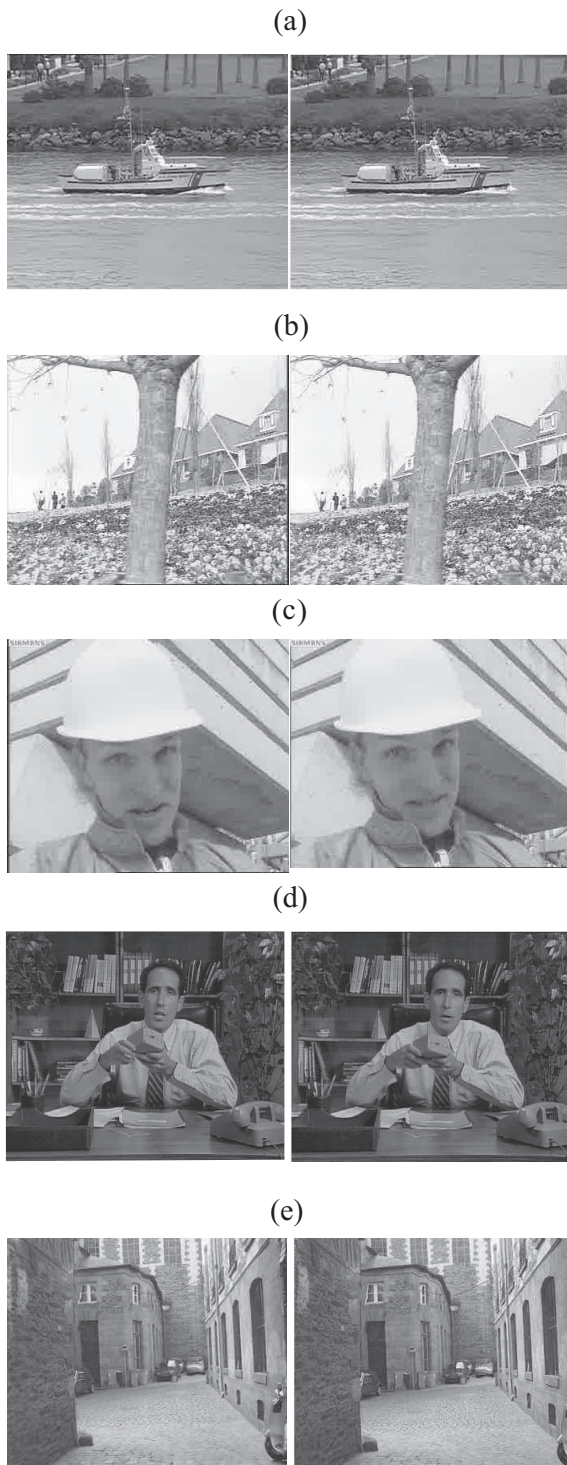


Figure 6 Anaglyph synthesized in case of use the Lucas & Kanade technique for video sequences: a) *Coastguard*, b) *Flowers*, c) *Foreman*, d) *Salesman* and e) *Alley*

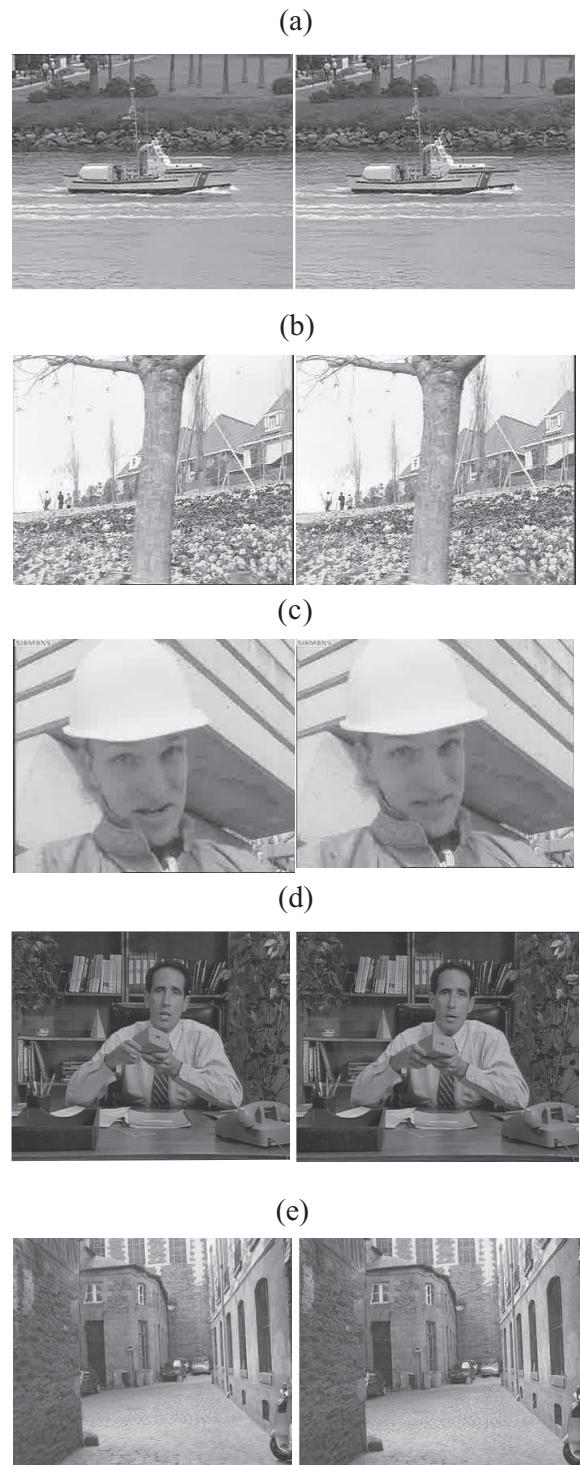


Figure 7 Anaglyph synthesized using the original Photoshop algorithm for video sequences: a) *Coastguard*, b) *Flowers*, c) *Foreman*, d) *Salesman* and e) *Alley*

It can be concluded analyzing this figure that PS algorithm for anaglyph construction allows sufficiently acceptable 3D perception in the video sequences, but more ghosting effects can be seen here in comparison with the proposed method results (see Figure 4). Additionally, we found that the time processing values for Photoshop algorithm are much more than in the case of the proposed algorithm usage.

Conclusions

In this paper, a novel method in the design of the high quality anaglyphs from color video sequences is introduced. The proposed method presents more precise motion information permitting better 3D perception. Anaglyphs are used as visualization technique viewing 3D information due to its simpler and cheaper way to display 3D information. Region based stereo matching calculation provides more reliable depth map in comparison with commonly used algorithms to obtain depth map like as Lucas & Kanade implementation.

Depth map interpolation for red component data allows the anaglyph construction and depth map compression via dynamic range computation permitting to realize better 3D perception due reduction the ghosting in anaglyphs, as it has been shown in the presented images. The proposed method has shown its efficiency in construction 3D color video sequences from 2D ones that are justified in color video sequences with different types of movements.

Acknowledgements

The authors thank National Polytechnic Institute of Mexico and CONACYT (grant 81599) for their support.

References

1. B. Blundell, A. Schwartz. *Volumetric three dimensional display systems*. Ed. Wiley. New York. Vol. 5. 2000. pp. 196-200.
2. M. Halle. "Autostereoscopic displays and computer graphics". *Comput Graph*. Vol. 31. 1997. pp. 58-62.
3. E. Dubois. "A projection method to generate anaglyph stereo". *IEEE International Conference on Acoustics, Speech, and Signal Processing*. Vol. 3. 2001. pp. 1661-1664.
4. I. Ideses, L. Yaroslavsky. "A method for generating 3D video from a single video stream". *Proc of the Vision, Modeling and Visualization*. Ed. Aka GmbH. Erlangen. Germany. 2002. pp. 435-438.
5. I. Ideses, L. Yaroslavsky. "New Methods to produce high quality color anaglyphs for 3-D visualization". *Lecture Notes in Computer Science*. Vol. 3212. 2004. pp. 273-380.
6. S. S. Beauchemin, J. L. Barron. "The computation of optical flow". *ACM Computing surveys*. Vol. 27. 1995. pp. 436-466.
7. J.L. Barron, D.J. Fleet, S.S. Beauchemin. "Performance of optical flow techniques". *International Journal of Computer Vision*. Vol. 12. 1994. pp. 43-77.
8. D. J. Heeger. "Model for the extraction of image flow". *Journal of the Optical Society of America*. Vol. 4. 1987. pp. 1455-1471.
9. D. J. Heeger. "Optical flow using spatiotemporal filters". *International Journal Computer Vision*. Vol. 1. 1988. pp. 279-302.
10. X. Huang, E. Dubois. "3D reconstruction based on a hybrid disparity estimation algorithm". *IEEE International Conference on Image Processing*. Vol. 8. 2006. pp. 1025-1028.
11. C. Zitnick, T. Kanade. "A cooperative algorithm for stereo matching and occlusion detection". *Robotics Institute Technical Reports*. No. CMU.RI-TR-99-35. Carnegie Mellon University. 1999.
12. H. H. Barker, T. O. Binford. "Depth from edge and intensity based stereo". *Proc. of the 7th International Joint Conference on Artificial Intelligence*. Vancouver. 1981. pp. 631-636.
13. D. Comaniciu, P. Meer. "Mean-shift: a robust approach toward feature space analysis". *IEEE Transactions Pattern Anal Machine Intelligence*. Vol. 24. 2002. pp. 603-619.
14. B. B. Alagoz. "Obtaining depth maps from color images by region based stereo matching algorithms". *OncuBilim Algorhythm and Systems Labs*. Vol. 08. 2008. Art.4. pp. 1-12.

15. I. Ideses, L. Yaroslavsky, B. Fishbain. "Real time 2D to 3D video conversion". *J. Real Time Image Proc.* Vol. 2. 2007. pp. 3-9.
16. I. Ideses, L. Yaroslavsky. "3 methods to improve quality of color anaglyph". *J. Optics A. Pure, Applied Optics.* Vol. 7. 2005. pp. 755-762.
17. I. Ideses, L. Yaroslavsky, B. Fishbain, R. Vistuch. "3D compressed from 2D video". *Stereoscopic displays and virtual reality systems XIV in Proc. of SPIE & IS&T Electronic Imaging.* Vol. 6490. 2007. pp. 64901C.
18. L. Yaroslavsky, J. Campos, M. Espínola, I. Ideses. "Redundancy of stereoscopic images: Experimental evaluation". *Optics Express.* Vol. 13. 2005. pp. 0895-10907.
19. L. Yaroslavsky, *Holography and digital image processing.* Ed. Kluwer Academic Publishers. Boston. 2004. pp. 600.
20. W. Sanders, D. McAllister. *Producing anaglyphs form synthetic images.* Conf. 5005: A stereoscopic displays and applications XIV. Proc. SPIE/IS&T. Vol. 5006. 2003. pp. 348-358.