Vision, Video and Graphics (2005)
E. Trucco, M. Chantler (Editors)

# Sensitivity Analysis in Image Synthesis

M. Trujillo and E. Izquierdo

Multimedia and Vision Lab, Electronic Engineering Department
Queen Mary, University of London, UK

**Abstract**

*Intermediate virtual images are used in the evaluation of disparity estimations. The analysis is based on the effect of disparity inaccuracies in the relative quality of virtual images. The peak-signal-to-noise ratio and the percentage of visual errors are used to assess the quality of virtual images generated with distorted disparity fields. Computer simulation results show that the peak-signal-to-noise ratio is more affected by a small magnitude of perturbations – two or three pixels – than by a high frequency of perturbations –15% or 20%. However, the percentage of visual errors is more affected by the frequency of perturbations. By using distorted disparity fields, the deterioration in the quality of virtual images is imperceptible to the human eye.*

I.4.8 [Scene Analysis]: Stereo

## 1. Introduction

In this paper, image synthesis is used for the evaluation of disparity estimations. Disparity estimation is still an open stereo vision problem. It is an inverse problem defined by an ill-posed functional. This means that the ideal disparity cannot be estimated. Image synthesis is based on a source description, which contains the recorded information about the objects in the world and their interactions, such as: shapes of the objects, properties of the objects' surfaces, lighting, and shading. Source descriptions commonly used are: scene geometry, the texture map, and a reflection model [BB78]. In this application, a disparity field is used as scene geometry description and a virtual image is created by image interpolation.

Two early works about image interpolation were those of Ullman and Basri [UB91] and Chen and Williams [CW93]. The former suggested that intermediate images can be created as a linear combination from the other images of the same scene, assuming an affine camera transformation. The latter proposed an image interpolation strategy using a transformation matrix to establish correspondences between pixels and linear interpolation to generate an intermediate image. They claimed that linear interpolation produces an exact solution when the optical axes of both views are parallel or the images have been rectified. Moreover, they showed that the linear interpolation produces the same results as perspective transformation when the optical axes of both views are parallel.

Regarding inaccurate disparity fields, Scharstein [Sch96] proposed a method for image synthesis from real images that is based on the scene geometry represented by disparity fields. He concluded that "*it is possible to efficiently synthesize realistic new views even from inaccurate and incomplete depth information*". In the same way, Izquierdo [Izq02] postulated that "*in many cases the effect of disparity* [in]*accuracy in the quality of virtual views is almost imperceptible*". Isgrò, Trucco, Kauff, and Schreer [ITK*04] discussed problems associated with 3-D image processing in the context of immersive telecommunications. They remarked on the need for highly accurate disparity fields in creating convincing virtual images.

Immersive telecommunications is one of applications of image synthesis. Tele-teaching, tele-conferencing and/or tele-collaboration approaches can be found in Schreer, Atzpadin, Askar, and Kauff [SAA*003] and Kauff and Schreer [KS02]. However, Smolic and Kauff [SK05] remarked: "*there are no original data to assess algorithms that compute intermediate views*". The evaluation of algorithms that compute intermediate images has to be in terms of efficiency and visual quality. Regarding visual quality, O'Sullivan, Dingliana, Giang and Kaiser [ODG*03] proposed a visual fidelity metric, which is based on three dynamic distortions. Teo and Heeger [TG94] studied the pattern sensitivity problem for perceptual distortion.

In this paper, a measure to approximate visual errors is used in order to establish differences in luminance that the human eye can perceive. The effect of disparity inaccuracies is examined by using ground truth stereo

images. The analysis is based on virtual images created with simulated distorted disparity fields and estimated disparity fields.

## 2. Disparity Estimation

Let $D_l$ and $D_r$ be the left and right images. Moreover, let $I_l : D_l \to G$ and $I_r : D_r \to G$ be the real positive bounded function mapping the discrete grid $D$ into $G = \{0,1,2,3,...\}$ representing the intensity maps. The image intensity at the point $(u,v)$ in the image $D_i$ is $I_i(u,v)$, $i = l,r$. Two points $p_l = (u_l, v_l)$ and $p_r = (u_r, v_r)$ in the images $D_l$ and $D_r$ are corresponding points if they are discrete image plane projections of the same real world surface point. In this case, the disparity vector $d_l$ in the direction from $p_l$ to $p_r$ is given by $d_l = (u_l - u_r, v_l - v_r)$. The disparity field from left to right is a mapping that assigns one disparity vector to each location on the image grid.

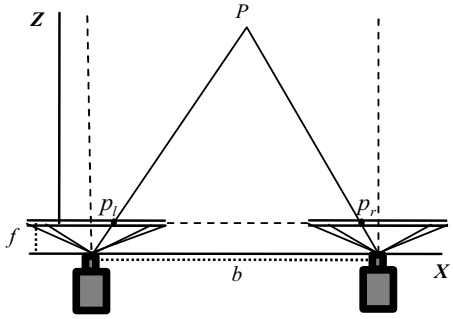



**Figure 1:** *Simplified stereo ring set-up with optical axes of both views parallel*

Disparity is inversely associated with depth. Assuming that either the optical axes of both views are parallel or the images have been rectified, in Figure 1, the disparity vector is $d = (d_u, d_v) = (u_l - u_r, 0)$. The relation between disparity and depth is:

$$d_u = f\,b \,/\, z\,k_u \,, \tag{1}$$

where $f$ is the focal distance, $b$ is the camera baseline, $z$ is the depth of a fixed point and $k_u$ is the horizontal pixel scale factor.

## 3. Intermediate Image Interpolation

Creating an arbitrary image requires the assignment of disparities with respect to the desired intermediate position. This process is carried out by projecting the calculated disparity vectors onto the intermediate image plane followed by a linear interpolation of the intensity values.

Disparity analysis is used in order to identify local characteristics of the disparity field that have implications for the optimum interpolation strategy [CW93] and [OGH*98].

### 3.1 Disparity analysis

Local characteristics of the disparity field are analyzed and used in the interpolation strategy. Assuming the camera set-up in Figure 1, three areas can be identified:

A1. Normal areas where the area is visible from both views.

A2. Right or Left occlusion areas where the area is invisible from one of the views (right/left).

A3. Right or Left contraction areas where the visibility resolution of the area from one view (right or left) is much higher than the other one, these areas being visible from both views.

### 3.2. Intermediate image synthesis

The disparity $d_u$, in Equation 1, changes proportionally to the camera base line $b$. If a virtual view position with distance $b'$ to one of the existing views is regarded, the disparity shift of the point must be scaled by the factor

$$s = \frac{b'}{b}. \tag{2}$$

By defining the position of the original left camera as reference, any virtual view position in between would have values $(0 < b' < b,\ 0 < s < 1)$.

In order to generate a pixel $p' = p + s \cdot d_u$ in a virtual intermediate image, the disparity vector has to be determined. If multiple vectors cross the area of the pixel, one of these should be selected. If no vector crosses the area of the pixel, linear interpolation has to be performed to resolve the missing reference. After selecting the disparity vector $d_u$ for use at the intermediate position $p_s = (u_s, v_s)$, it is possible to determine which intensity values from the left and right original images have to be referenced by this vector, and to use a weighted average for creating the intensity value $I_s(p_s)$ of the virtual image $p_s$ in the intermediate plane:

$$I_s(p_s) = w_l \cdot I_l(u_s - s \cdot d_u, v_s) + w_r \cdot I_r(u_s + (1-s) \cdot d_u, v_s) \,. \tag{3}$$

In order to define the two weighting factors, $w_l$ and $w_r$ performing a convex combination, a constraint has to be imposed: $w_l + w_r = 1$. The two weighting factors have to be defined according to the virtual position in order to make possible the reconstruction of the original position by linear interpolation. Consequently, it is necessary to take into account the position parameter $s \in [0,1]$ and local characteristics of the disparity field. An adequate strategy for interpolation is:

- In normal areas, the weighting factor can be settled as $w_l = s$ and $w_r = (1-s)$.

a) Tsukuba


b) Venus

c) Map

**Figure 2:** *Left images, ground-truth disparity fields and occlusion areas (black) and areas of discontinuity (white) from left to right.*

- In the left or right contraction areas, apply more weight to the view which is at the spread side of the disparity.
- In the left (or right) occlusion areas, $w_l = 1$ or $w_r = 0$ in order to use information exclusive to the view where the area is visible.

## 4. Sensitivity Analysis

Sensitivity analysis is a procedure for determining the sensitivity of the output to changes in the input. Our objective is to analyse the effect of disparity inaccuracies (input) in the relative quality of virtual images (output).

The evaluation of the quality of virtual images generated with distorted disparity fields is based on the peak-signal-to-noise ratio (PSNR) and the percentage of visual errors.

Let $I_{true}$ and $I_{estimated}$ be the intensity maps of the intermediate images synthesized using the ground-truth disparity field and an estimation of the disparity field, respectively.

### 4.1. Peak –signal-to-noise ratio

The PSNR is calculated on a region of interest *R*. For an 8-bit (0-255) image, the PSNR (in dB) is defined as:

$$PSNR = 10\log_{10}\frac{255^2}{MSE}, \quad (4)$$

the mean-square error (MSE) is given by

$$MSE = \frac{1}{\#(R)}\sum_{i\in R}\sum_{j\in R}[I_{true}(i,j) - I_{estimated}(i,j)]^2, \quad (5)$$

where #(*R*) is the cardinality of a region *R*.

### 4.2. Visual errors

In order to quantify visual errors, the difference in intensity values

$$e_{i,j} = \begin{cases} 1 & if \ \left| I_{true}(u_i, v_i) - I_{estimated}(u_i, v_i) \right| > \delta_e \\ 0 & if \ \left| I_{true}(u_i, v_i) - I_{estimated}(u_i, v_i) \right| \leq \delta_e \end{cases} \quad (6)$$

is used. We expect that this error measure approximates the subjective error perception of the human visual system.

The threshold $\delta_e$ is defined using the sensitivity curve in Figure 3.

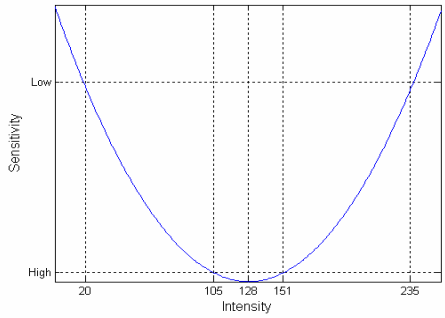



**Figure 3:** Sensitivity curve

Sensitivity curve is used in compression of digital images [VF96]. The behaviour of the sensitivity curve depends on the intensity values. Small changes of intensity values in [105, 151] are perceived by the human visual system. The human visual system is robust against small changes of intensity values in [0, 20] ∪ [235, 255]. In line with this behaviour, three areas are established in the sensitivity curve: high, medium, and low. Intensity values in [105, 151] belong to the high area. Intensity values in [21, 104] ∪[152, 234] belong to the medium area. Intensity values in [0, 20] ∪ [235, 255] belong to the low area. The threshold is defined according to

$$\delta_e = \begin{cases} 2 & if \ I_{true}(u_i, v_i) \in high\ area \\ 10 & if \ I_{true}(u_i, v_i) \in medium\ area \\ 20 & if \ I_{true}(u_i, v_i) \in low\ area \end{cases} \quad (6.a)$$

## 5. Results of Computer Simulations

Experiments are designed and evaluated using ground-truth images available at the Middlebury Stereo Vision web page [MSV]. Figure 2 shows the test stereo images, ground-truth disparity fields, and occlusion areas and areas of discontinuity. Tsukuba comprises four objects in the foreground and many details in the background; Venus contains large textureless areas; and Map consists mainly of texture.

Two factors are considered: the frequency and the magnitude of perturbations. The frequency is the percentage of pixels modified in a specific region and the magnitude is the value (pixels) added to the truth-disparity value to produce a distorted disparity.

According to Shcarstein and Szeliski's performance measures [SS01], the frequency of errors in disparity estimations is less than 20% in depth discontinuity image regions and less than 15% in low-textured areas depending on the algorithm used. Consequently, four levels of the frequency of perturbations are used: 5%, 10%, 15%, and 20%. The levels of the magnitude of perturbations are bounded by the range of disparity. A single experiment is the combination of one level of frequency with one level of magnitude. Each experiment is replicated 30 times.

Firstly, perturbations in areas of discontinuity are analyzed. Friedman's test is used to determine whether differences obtained among PSNR values are due to random variations or to the effect of perturbations Secondly, textureless areas are analyzed. Finally, intermediate virtual images are created by using estimated disparity fields.

In computer simulations, the calculation of the PSNR and the percentage of visual errors are based on regions rather than the whole image. The region corresponds to the perturbed area.

### 5.1. Areas of discontinuity

Figures 4 to 6 show means – over 30 replicates – of PSNR values obtained for different levels of magnitude and frequency of perturbations.

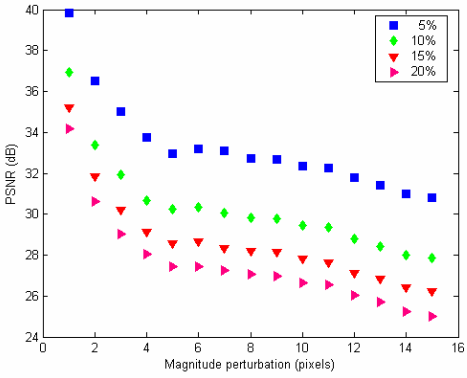



**Figure 4:** *Means of PSNR values for Tsukuba*

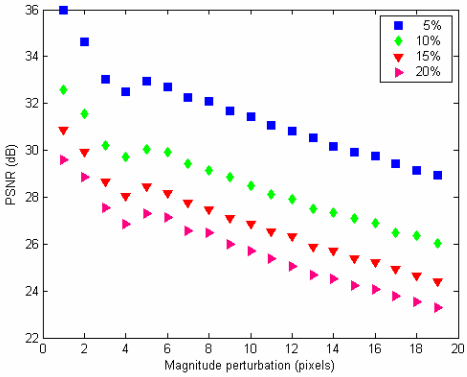



**Figure 5:** *Means of PSNR values for Venus*

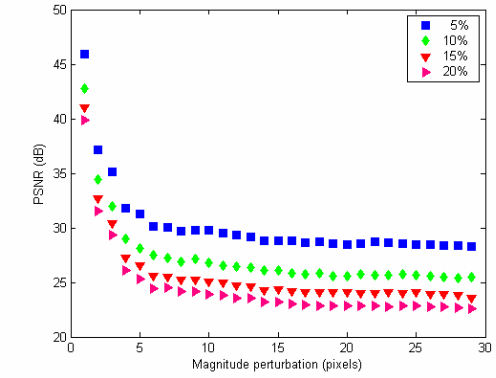



**Figure 6:** *Means of PSNR values for Map*

Results obtained for Tsukuba, in Figure 4, show that increasing the magnitude of perturbations from one pixel to two pixels produced a significant reduction in the PSNR values. This means that the relative quality of an intermediate virtual image was affected by this change. However, after four pixels the reduction in PSNR values was smoother. PSNR values obtained for Venus, in Figure 5, were lower than values obtained for Tsukuba. For Map in Figure 6, PSNR values decreased exponentially.

The PSNR values behaved according to the image composition. Map – consists mainly of texture – was more affected by the perturbations than Tsukuba and Venus.

Figures 7 to 9 show means – over 30 replicates – of percentages of visual errors obtained for different levels of magnitude and frequency of perturbations.

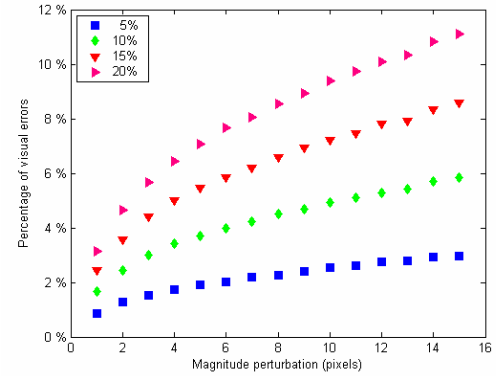



**Figure 7:** *Percentage of visual error for Tsukuba*

When the magnitude of the perturbations was equal to one, there were small differences among the percentages of visual errors from different levels of frequency. However, there were differences among percentages of visual errors of different levels of frequency when the magnitude of perturbation was equal to the maximum of the disparity range. This can be observed in results obtained for Tsukuba, in Figure 7. Percentages of visual errors were less than 4% when the frequency of perturbations was equal to 5%. However, when the frequency was equal to 20%, percentages of visual errors increased from 2.94% –one pixel perturbation – to 11.37% – 15 pixels perturbations. Similar results were obtained for Venus and Map.

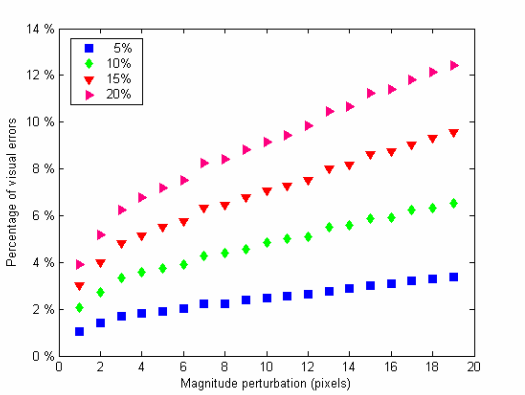



**Figure 8:** *Percentage of visual error for Venus*

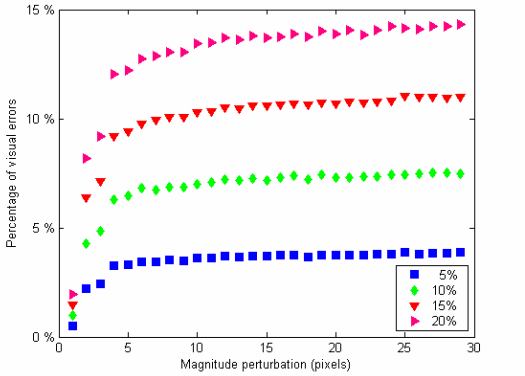



**Figure 9:** *Percentage of visual erro for Map*

The percentage of visual errors was more affected by the frequency of perturbations than by the magnitude of perturbations.

In Tables 1 to 3 the maximum corresponds to a PSNR value obtained when the magnitude of the perturbations was equal to one pixel and the minimum corresponds to a PSNR value obtained when the magnitude of the perturbations was equal to the maximum of the disparity range.

**Table 1**: *Statistics of PSNR values for Tsukuba*

| Freq. | 5% | 10% | 15% | 20% |
|---|---|---|---|---|
| max | 41.95 | 38.35 | 36.12 | 34.83 |
| mean | 33.29 | 30.34 | 28.68 | 27.54 |
| min | 30.10 | 27.26 | 26.06 | 24.69 |

**Table 2:** *Statistics of PSNR values for Venus*

| Freq. | 5% | 10% | 15% | 20% |
|---|---|---|---|---|
| max | 38.58 | 33.69 | 32.11 | 30.21 |
| mean | 31.53 | 28.62 | 26.96 | 25.82 |
| min | 28.36 | 25.64 | 24.05 | 23.12 |

**Table 3:** *Statistics of PSNR values for Map*

| Freq. | 5% | 10% | 15% | 20% |
|---|---|---|---|---|
| max | 47.72 | 44.30 | 42.35 | 40.96 |
| mean | 30.21 | 27.31 | 25.61 | 24.49 |
| min | 27.22 | 24.58 | 23.21 | 22.10 |

For Tsukuba and Venus the difference between maximum and minimum PSNR values was up to 10 dB while for Map this difference was up to 20 dB. The difference between the 20% and 5% frequency of perturbations was up to 7 dB for the three test images. Regarding the frequency of perturbations, the largest reductions in PSNR values were obtained when the frequency changed from 5% to 10%.

The PSNR value, as a measure of the relative quality of an intermediate image, was more affected by the magnitude of the perturbations than by the frequency.

Tables 4 to 6 present statistics of percentages of visual errors for Tsukuba, Venus, and Map, respectively The largest percentages of visual errors were obtained with Map when the frequency of perturbations was 20% and the magnitude of the perturbations was the maximum of the disparity range (29 pixels).

**Table 4***: Statistics of percentages of visual errors for Tsukuba*

| Freq. | 5% | 10% | 15% | 20% |
|---|---|---|---|---|
| max | 3.17% | 6.00% | 8.74% | 11.37% |
| mean | 2.18% | 4.26% | 6.24% | 8.11% |
| min | 0.71% | 1.52% | 2.27% | 2.94% |

**Table 5***: Statistics of percentages of visual errors for Venus*

| Freq. | 5% | 10% | 15% | 20% |
|---|---|---|---|---|
| max | 3.63% | 6.78% | 9.80% | 12.75% |
| mean | 2.42% | 4.72% | 6.90% | 8.99% |
| min | 0.87% | 1.81% | 2.71% | 3.64% |

**Table 6:** *Statistics of percentages of visual errors for Map*

| Freq. | 5% | 10% | 15% | 20% |
|---|---|---|---|---|
| max | 4.26% | 7.93% | 11.58% | 14.98% |
| mean | 3.47% | 6.78% | 9.91% | 12.88% |
| min | 0.39% | 0.74% | 1.22% | 1.70% |

Percentages of visual errors for Map, in Table 6, were relatively small – up to 1.7% – when the magnitude of perturbations was one pixel.

Friedman's test was used to determine whether differences obtained among PSNR values were due to random variations or to the effect of perturbations. Table 7 shows results of Friedman's test. We can conclude that the differences obtained among PSNR values were due to the effect of perturbations.

**Table 7:** *Friedman's test*

| Image pairs | Frequency | | Magnitude | |
|---|---|---|---|---|
| | $?^2$ | Pr > $?^2$ | $?^2$ | Pr > $?^2$ |
| Tsukuba | 407.10 | 0 | 404.61 | 0 |
| Venus | 437.18 | 0 | 341.28 | 0 |
| Map | 435.97 | 0 | 338.67 | 0 |

### 5.2. Textureless areas

Secondly, textureless areas were analyzed. The relation between the magnitude of perturbations and PSNR values, for the four levels of frequency, was similar to the relation established in areas of discontinuity. However, PSNR values were less affected by perturbations in textureless areas. Table 8 show the minimum of PSNR values obtained for different levels of frequency.

**Table 8:** *Minimun of PSNR values*

| Freq. | 5% | 10% | 15% | 20% |
|---|---|---|---|---|
| Tsukuba | 36.88 | 34.80 | 33.42 | 32.24 |
| Venus | 33.92 | 31.42 | 30.15 | 28.92 |
| Map | 27.39 | 25.39 | 23.97 | 22.83 |

Table 9 contains the maximum of percentages of visual errors obtained in textureless areas. Map contained the largest percentages while Venus had lower percentages of visual errors.

**Table 9:** *Maximum of percentages of visual errors*

| Freq. | 5% | 10% | 15% | 20% |
|---|---|---|---|---|
| Tsukuba | 1.03% | 1.96% | 2.82% | 3.71% |
| Venus | 0.94% | 1.84% | 2.54% | 3.30% |
| Map | 4.17% | 7.43% | 10.14% | 12.68% |

### 5.3. Intermediate virtual images using estimated disparity fields

Finally, intermediate virtual images were created by using estimated disparity fields. These disparity fields were calculated using a bidirectional local maximization algorithm and the sum of absolute differences (SAD) using a 7x7 block size. Linear interpolation was used for estimating unmatched points in order to obtain dense-disparity fields.

In order to visualize errors, an error map is generated using

$$E(i,j) = 2\left[I_{true}(i,j) - I_{estimated}(i,j)\right] + 128. \quad (7)$$

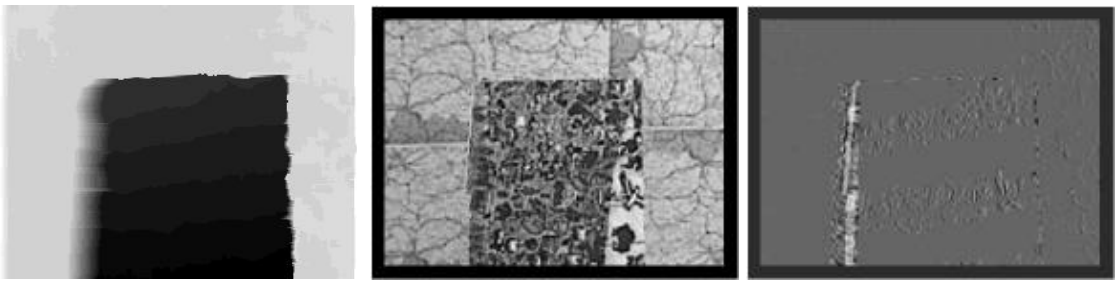
Since the differences between the true and the estimated intensity value can be either positive or negative, the factor 128 is added.

A perfect image synthesis will produce an error map represented by a uniform field with value 128.

Figure 10 shows estimated disparity fields, intermediate virtual images, and error maps.

a) Tsukuba

b) Venus



c) Map

**Figure 10:** *Distorted disparity field, virtual intermediate images and error maps (from left to right)*

Intermediate virtual images calculated by using distorted disparity fields are perceived by the human eye as realistic. However, error maps show areas that contain differences in intensity values. By looking at Map's intermediate virtual image in Figure 10, we can observe visual artefacts close to the occlusion area. The error map reveals the largest differences in intensity values in this area.

**Table 10:** *PSNR, MSE, and percentage of visual errors for virtual images using distorted disparity fields*

| Image pairs | PSNR (dB) | MSE | Visual errors |
|---|---|---|---|
| Tsukuba | 30.28 | 2.45E+03 | 2.95% |
| Venus | 29.66 | 3.34E+03 | 8.69% |
| Map | 25.88 | 3.08E+03 | 15.30% |

The PSNR, the MSE and the percentage of visual errors were calculated using the whole image. Table 10 present PSNR values, MSE values, and percentages of visual errors. The largest percentage of visual errors was obtained for Map, 15.30%.

## 6. Conclusions

A visual error measure was used in order to count the number of pixels with intensity differences perceived by the human eye.

Two factors – frequency and magnitude of perturbations – were considered in the sensitivity analysis. Discontinuity and textureless areas were analyzed. In these two areas, PSNR values decreased exponentially. Percentages of visual errors were more affected by the

frequency of perturbations than by the magnitude of perturbations

Based on the simulation results, small perturbations – one, two or three pixels – produced a large reduction in the PSNR value. Visual errors appeared to be robust against small perturbations.

Although there was deterioration in the quality of virtual images that were created by using estimated disparity fields, these new intermediate images are perceived by the human eye as realistic.

**Acknowledgement**

Maria Trujillo, assistant professor at University of Valle, Cali-Colombia, acknowledges the financial support provided by the University of Valle to carry out this research.

**References**

**BB78** BADLER N., BAJCSY R.: Three-dimensional representation for computer graphics and computer vision. In *ACM SIGGRAPH* 1978, pp. 153-160.

**CW93** CHEN S. E., WILLIAMS L.: View Interpolation for Image Synthesis. In *ACM SIGGRAPH*, 1993, pp. 279-288.

**ITK*04** ISGRO, F.; TRUCCO E., KAUFF P.; SCHREER E.*:* Three-dimensional image processing in the future of immersive media. In IEEE *Transactions on Circuits and Systems for Video Technology*, Vol. 14 , No. 3 , 2004, pp. 288 – 303.

**Izq02** IZQUIERDO E.: How accurate should be disparity estimated for image-based rendering? *In International Symposium on Video/Image Processing and Multimedia Communications*, Zadar, Croatia, 2002, pp. 69-74.

**KS02** KAUFF P., SCHREER O.: An immersive 3D video-conferencing system using shared virtual team user environments. In *ACM-CVE'02*, Bonn, Germany, 2002, pp. 105-112.

**ODG*03** O'SULLIVAN **C.,** DINGLIANA J**.,** GIANG T**.,** KAISER M. K.: Evaluating the visual fidelity of physically based animations. In *ACM* 2003, pp. 527-536**.**

**OGH*98** OHM J. R., GRÜNEBERG K.. HENDRIKS E., IZQUIERDO M. E., KALIVAS D., KARL M., PAPADIMATOS D., REDERT A.: A Realtime Hardware System for Stereoscopic Videoconferencing with Viewpoint Adaptation. In *Signal Processing Image Communication*, Vol. 14, 1998, pp. 147-171.

**Sch96** SCHARSTEIN D.: Stereo Vision for View Synthesis. In IEEE *Computer Society Conference on Computer Vision and Pattern Recognition*, 1996, pp. 852-858.

**SS01** SCHARSTEIN D., SZELISKI R.: A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. In *Microsoft Corporation, Technical Report MSR-TR-2001-81*, 2001, 59 pages.

**SAA*03** SCHREER O., ATZPADIN N., ASKAR S., KAUFF P.: Advanced 3D Signal Processing for Virtual Team User Environments. In IEEE *International Conference on Multimedia and Expo*, Vol. 2 , 2003, pp. 61-64.

**SK05** SMOLIC A., KAUFF P.: Interactive 3-D Video Representation and Coding Technologies. In *Proceeding of the IEEE*, Vol. 93, No. 1, 2005, pp. 98-110.

**TG94** TEO P. C., HEEGER D. J.: Perceptual image distortion. In Proceedings of *SPIE Human Vision, Visual Processing and Digital Display*, Vol. 2179, 1994, pp. pp. 127-141.

**UB91** ULLMAN S., BASRI R. Recognition by linear combination of models. IEEE *Transaction on Pattern Analysis and Machine Intelligence*, Vol. 13, No. 10, 1999, pp. 992-1006.

**VF96** VAN DEN BRADEN CH. J., FARREL J. E.: Perceptual quality metric for digitally coded color images. In Proceedings of the *European Signal Processing Conference*. Trieste, Italy, 1996, pp. 1175-1178.

**WHS95** WERNER T., HERSCH R. D., SLAVÁC V.: Rendering Real-World Objects Using View Interpolation. In IEEE *International Conference in Computer Vision*, 1995, pp. 957-962.

**WS82** WYSZECKI G., STILES W. S.: Color Science. John Wiley and Sons, 1982.

**MSV** http://www.middlebury.edu/stereo