A computer aided visual model
for ensuring video watermarking transparency

Jana Dittmann, Anirban Mukherjee and Martin Steinebach[1]

GMD - German National Research Center for Information Technology, Darmstadt, Germany

ABSTRACT

Digital watermarking is a technology for copyright protection and protection against unauthorized access and modification of multimedia material. The most important properties of digital watermarking techniques are robustness, security, imperceptibility/ transparency, complexity, capacity and possibility of verification. Robustness means resistance to "blind", non-targeted modifications, or common media operations. A transparent watermark causes no artifacts or quality loss. A maximum of robustness cannot be achieved at the same time as a maximum of transparency as a higher robustness requires stronger media modifications.

Transparency is based on the properties of the human visual system or the human auditory system. It is an often-neglected part of a watermarking evaluation scheme. We introduce a computer aided visual model based on a visual modulation threshold function which is used to test the degree of transparency of the watermark in watermarked multimedia material using linear system theory. We describe our test environment for the model and show how it is implemented in 5 steps representing the essential parts of the visual model: Sampling, band pass contrast response, oriented response, transducer and distance. The model takes two digital images as the input of the and returns a probability that an observer can distinguish the two pictures.

Keywords: Watermarking, Gaussian and Laplacian Pyramid, Modulation threshold function, spatial filter, MPEG video

1.   BACKGROUND AND MOTIVATION

The advance of new technology in the field of computer networks makes it easy to transfer, access and modify multimedia material. Therefore the multimedia community feels a need to protect multimedia material from the attacks of unauthorized persons. Digital images and video can be copied without loss of its quality. Even though such copying violates the copyright law, it is widespread. Therefore, copyright protection of multimedia material is a challenging issue.

In this section we give a brief introduction to the basic concepts of this paper. In section 1 the properties of a digital watermark are discussed. Transparency, one of the most important watermarking parameters is explained, and ways of evaluating transparency are discussed, introducing a model of human perception in section 2. In section 3 this model is implemented in a watermarking environment. Section 4 provides test results for the model and in section 5 open problems are discussed. An example application for the model is described in section 6. Section 7 gives a brief conclusion.

1.1 Digital Watermarking

Digital watermarking is a technology for copyright protection and protection against unauthorized access and modification of multimedia material. Robust digital watermarking can be used to claim copyright protection by embedding authors or producers information. Fragile watermarking technology addresses the recognition of manipulations. Today watermarking concentrates on one type of media like still images or audio. For multimedia protection, a combination of different types of media, we have introduced the idea of mutual watermarking in [5]. Different types of media are linked with watermarks referring to the other media. This is used in the context of content fragile watermarking [6] to offer a maximum of integrity protection.

---

[1] Correspondence: Email: martin.steinebach@darmstadt.gmd.de; WWW: http://www.darmstadt.gmd.de/~steineba

1.2 Properties of Digital Watermarks

Watermarks embedded in multimedia data for enforcing the copyright must recognize the data as the property of the copyright holder. They must be difficult to remove, even after various media transformation processes like conversion, compression, scaling, rotating, shearing etc. A robust watermark should always remain present in the data. Today existing watermarking technology has different security problems regarding robustness and visual artefacts. To prevent any copyright forgery, the key to copyright watermarking is to provide security and robustness of the embedded watermark against a variety of attacks, which include for example:

- Detecting the location of an embedded watermark by comparing different copies of watermarked material
- Altering the embedded watermark through visual analysis
- IBM attack: Instead of introducing a new watermark with an own algorithm and claiming the authorship, a counterfeit original of a watermarked picture is produced by subtracting a watermark from the marked media.

We have implemented the watermark embedding and extraction scheme described by Dittmann et al[1]. The scheme proposed in that paper is based on ideas of Jiri Fridrich. It is designed to overcome two main shortcomings of the original: retrieval without the original and embedding of binary coded labels. To ensure the last requirement instead of one overlaying pattern over the whole picture we add an 8 by 8 pattern over 8 by 8 block of every picture of the MPEG-II video.

1.3 Transparency and its Evaluation

Transparency is based on the properties of the human visual system or the human auditory system[5]. A transparent watermark causes no artefacts or quality loss. Though transparency is one of the most important quality parameter of watermark, it is an often neglected part of a watermarking evaluation scheme. After embedding the watermark, the author performs a subjective test to check the transparency.

Performing this type of subjective test has some inherent disadvantages. The main problem is cost. It is also a time-consuming process, especially when extensive training is necessary to adapt the experimental observer to inherent differences among the different multimedia material.

We apply a computer aided model for this discrimination task. In this model the output of different filters, tuned to different frequencies, spatial positions and orientations are passed through a sigmoid shaped non-linearity and then to a summation channel that converts the outputs of many channels to a single scalar value. Thereby a model of human vision is created, enabling computer based subjective tests without a human observer.

2. MONOCHROME VISUAL MODEL

To design a visual model, we use the knowledge of linear system theory[10]. Here, an optical system is treated as a two dimensional linear system that is linear in intensity and can be characterized by a two-dimensional transfer function [2].
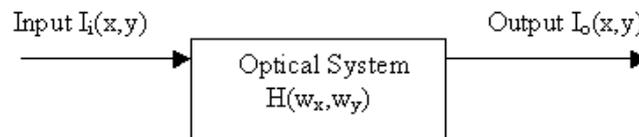

Figure 1: Linear system analysis of an optical system

If $\vartheta_i(w_x,w_y)$ and $\vartheta_o(w_x,w_y)$ are the frequency spectra of the input intensity distribution of the object $I_i(x,y)$ and output intensity distribution of the image $I_o(x,y)$ then the modulation transfer function of the optical system is defined as

$$M(w_x,w_y) = | \vartheta_o(w_x,w_y) | / | \vartheta_i(w_x,w_y) | \qquad (1)$$

The visual modulation threshold function Mt(u) gives the modulation depth threshold of a just observable sine wave bar pattern as a function of the spatial function u on retina of the eye. Often the reciprocal value of the modulation threshold

function is used. This function is called the contrast sensitivity function, the contrast below which grating is invisible: $S(u) = 1/M_t(u)$. $S(u)$ can be approximated by the formula

$$S(u) = 1/M_t(u) = au \exp(-bu)sqrt(1+c \exp(bu)) \tag{2}$$

where $a = 540(1+0.7/L)^{-0.2} / (1+12 / [w(1+u/3)^2])$, $b= 0.3(1+100 / L)^{0.15}$, $c= 0.6$ and L and u are display luminance in cd / m^2 and display width in degree respectively.

3. SOFTWARE SIMULATION

In this section, the computer aided implementation of the visual model is discussed. Section 3.1 describes our test environment for which the computer aided visual model is implemented. Section 3.2 shows how the visual model (figure 2)[8] is implemented in 5 steps representing the essential parts of the visual model.

3.1 Environment

The visual model discussed in this paper is used in an environment for a combined video and audio watermarking approach[5]. We use the MPEG 2 decoder, provided by MPEG Software Simulation Group, to decode the original MPEG video and the corresponding watermarked one. We randomly (with current time as a seed) decode the pictures from both MPEG videos to bypass the large time overhead involved in decoding of all the pictures. We use the portable picture map (ppm) format to store the picture. As we apply different types of watermarking algorithms, we need tools to extract media streams out of the system stream, convert MPEG data to image or PCM data and to rewrite the system stream with the changed data.

We have developed an algorithm for the extraction of audio and video elementary stream from the MPEG system stream and putting it back to the original one after embedding the watermark in the elementary streams. During extraction stage, we extract the packets for a particular elementary stream and apply the watermark embedding technique on the decoded version of the extracted elementary stream. We then replace these extracted and watermarked data packets in the original system stream.

3.2 Computer Aided Visual Model

A visual model is computer aided to the extent that components of the model have the same functional response as physiological mechanisms in the visual pathways of the brain. The basic model, described in this paper, was developed to quantify the visibility of the cockpit display information so that the display designers might access the effect that specific design choices would have on crew performance over a broad range of flight scenarios. In the following we discuss our implementation of the computer aided visual model for our watermarking environment.

Stimuli

The model takes two digital images, expressed as sample luminance distribution, as the input of the system. In addition, several observer parameters are given, including distance of the observer from the images, fixation depth and eccentricity of the images in the observer's visual field (expressed in degrees of visual angle). The model returns a probability that an observer can distinguish the two pictures.

In our watermarking application, several parameters of the original model, for example observer distance, are not important and can be ignored without influencing model reliability.

Step 1: Sampling

The image is re-sampled to model the sampling by the cones. For images with the eccentricity of $0°$, the image is re-sampled in a square grid of 120 pixels per degree to generate a modeled retinal image of 512 by 512 pixels, covering approximately $4.25°$ of visual space in each linear dimension. For eccentricity greater than $0°$, the retinal pixel density is not 120 pixels per degree but $120 / (1+ke)$, where e is the eccentricity and k is set to 0.4, the value estimated from psychophysical data by Burt and Aleson[4]. In our simulation, we assumed the eccentricity of the display to be $0°$.

Step 2: Bandpass contrast response

In the next step the raw luminance signal is converted to units of local contrast, described in [3]. First the image is decomposed into "Laplacian Pyramid". Let us suppose that $g0(i,j)$ is the original image and $g1(i,j)$ is the result of applying the low pass filter to $g0$. The prediction error is $L_0(i,j) = g_0(i,j) - g_1(i,j)$.
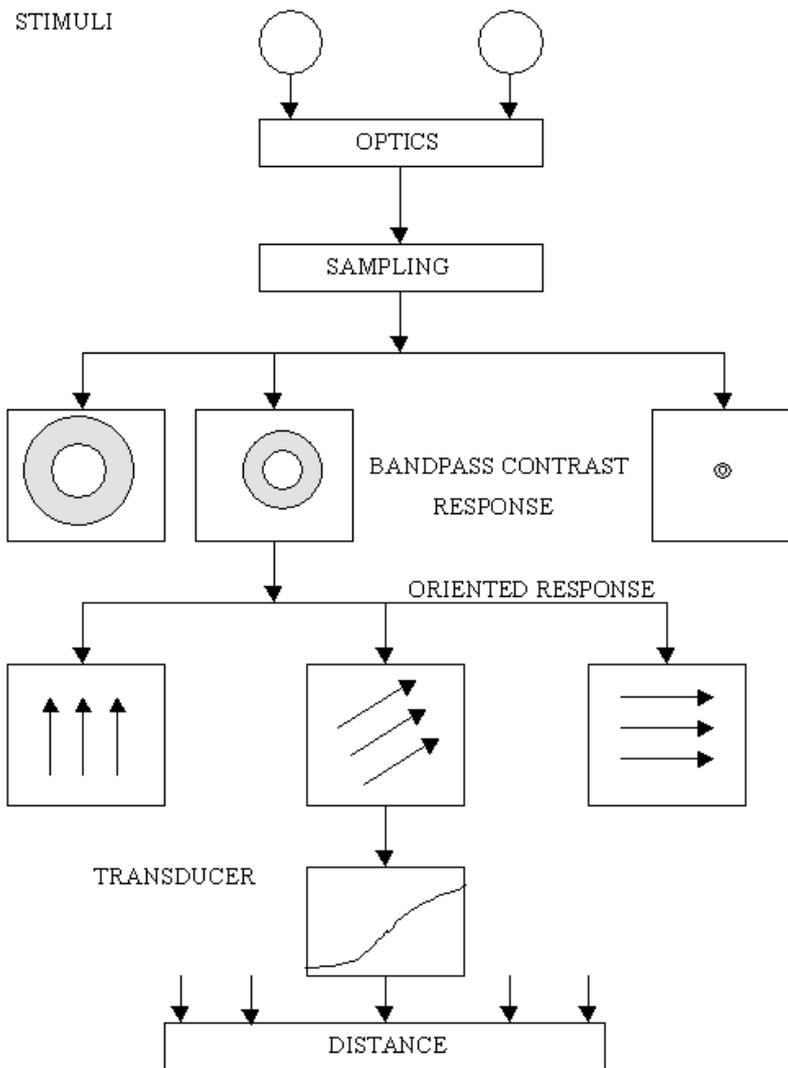
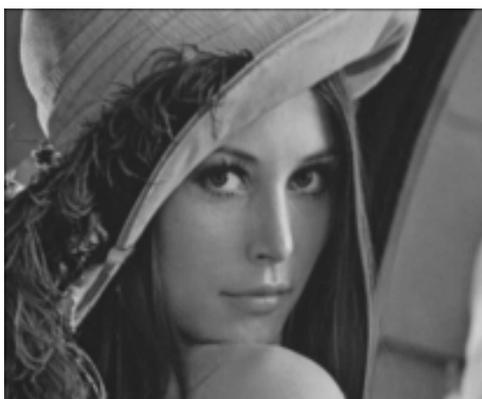**Figure 2 : Visual model scheme**

Figure 3: Lena g0(i,j)

Figure 4: Low pass version of Lena g1(i,j)

As $g_1(i,j)$ is the low pass version of the original image $g_0(i,j)$, it can expressed with a reduced sample rate. The reduced image is itself low pass filtered to produce $g_2(i,j)$ and a second error image is obtained:

$$L_1(i,j) = g_1(i,j) - g_2(i,j) \qquad\qquad (3)$$

By repeating these steps several times a sequence of error images L1, L2,...., Ln is created. In this implementation each image is smaller than its immediate predecessor by a factor of two due to reduced sample density. If we now consider the error image placed one over another, it looks like a tapered pyramid data structure. The value at each node is the difference between two gaussian-like or related functions convoluted with the original image. The difference between these two functions is termed as laplacian function.

Figure 5: Difference picture L0(i,j) (contrast enhanced)

We have generated a laplacian pyramid of seven steps resulting in seven bandpass levels with peak frequencies from 32 to 0.5 cycles per degree, each level separated from its neighbor by one octave. If we divide the laplacian image by the corresponding gaussian image, upsampled by a factor of two, we will get a local measure of contrast, which is roughly approximated by the Weber ratio.

Step 3: Oriented response

In the following stage, we rotate the image in three different orientations. The grating contrast detection threshold in each orientation and each pyramid level normalizes the luminance value.

Step 4: Transducer

Next in the transducer stage, each normalized luminance value is passed through a sigmoid type non-linearity function of the form

$$T(e_i) = (k+2)|l_i^{\ n}| \ / \ k|l_i|^{(n-w)} + |l_i|^m + 1 \qquad (4)$$

where n = 1.5, m = 1.1, w = 0.068 and k = 0.1.
This nonlinearity function T is required to reproduce the dipper shape of the contrast discrimination function.

Step 5: Distance

After the transducer stage, each image's model output for each spatial position can be described as a m dimensional vector, where m (21) is the number of frequency levels (7) times the number of orientations (3). In the box, marked as "Distance" in figure 2, distance between two vectors is calculated in the following way

$$D_j = \left(\max \left[\ \sum_k |T_{j,k}(s_1) - T_{j,k}(s_2)|Q \ \right] \ \right)^{1/Q} \qquad (5)$$

where j indexes over spatial position, k indexes over 21 frequency times orientations, s_1 and s_2 are two input images and Q is a parameter set to 2.4.

Figure 6: Original Lena

Figure 7: Watermarked Lena

The result of this stage is the distance value between two pictures, original (figure 6) and watermarked (figure 7). We have calculated this distance between several test pictures and depending on the result of this subjective test, we have set the decision threshold. We have decoded pictures from MPEG movie clips to ppm (portable picture map) and compressed it in JPEG format with the quality from 10 to 90 in steps of 10. The test results are discussed in section 4, in the appendix test result data tables for five images can be found.

The quality lets the user trade off compressed file size against image quality: the higher the quality setting, the larger the JPEG file, and the closer the output image will be to the original input. The quality 100 will generate a quantization table of

all 1's, minimizing loss in the quantization step (but there is still information loss in sub-sampling, as well as round-off error). Quality values below 50 will produce very small files of low image quality.

4. TESTRESULTS

We have tested our model with 5 different pictures. Lena is a photograph. The others are decoded from MPEG-1 movie clips. From the tables 2-5, given in the appendix, it is concluded that we can not make a decision about quality degradation on the basis of RMS (**R**oot **M**ean **S**quare) difference between two pictures. In these tables, the inference Y means difference between original and JPEG compressed image perceived by 70% or more observers. The inference Y/N means difference perceived by 40-60% observers and inference N means difference not perceived by 70% or more observers. If we look at the five tables, the inference Y/N occurs at JND (**J**ust **N**oticeable **D**ifference) values 4 to 5. But in that range, the RMS difference varies from 1.43228 to 7.27618 (pixel value is normalized between [0,1]). As small changes in the high frequency component of an image are not noticed by human observers, RMS measure is not a good estimate for image quality. The percentage size given in table 2-5 is calculated with respect to the file size of the JPEG compressed image with 100% quality.

We have kept six different stages of decisions:

Jndvalue	Quality
≤ 4.20	6: No difference observed
> 4.20 to 5.00	5: Easy differences, not disturbing
> 5.00 to 5.50	4: Differences well ascertainable, easily disturbing
> 5.50 to 6.00	3: Differences well ascertainable, strongly disturbing
> 6.00 to 6.50	2: Large differences, even still acceptable
> 6.50	1: Strong differences, no longer acceptable

Table 1: JND values and quality

5. OPEN PROBLEMS

We have set a hard threshold to decide about the quality degradation due to embedding of watermark. But the threshold should be adaptive in nature and dependent on the picture to be tested. If the image is inherently edgy in nature, the first artifacts are observed near the edge. If the background of the image is absolutely dark, the model fails to predict the difference between original and JPEG compressed image. In picture marsobserver (figure 8 and 9) a human observer will not perceive a difference between original image and JPEG compressed one, but the model gives the JND value of 43.3721, which is abnormal. It is also observed that in all the pictures except marsobserver, the JND value decreases with increasing quality, but in the case of marsobserver almost random changes seem to occur.
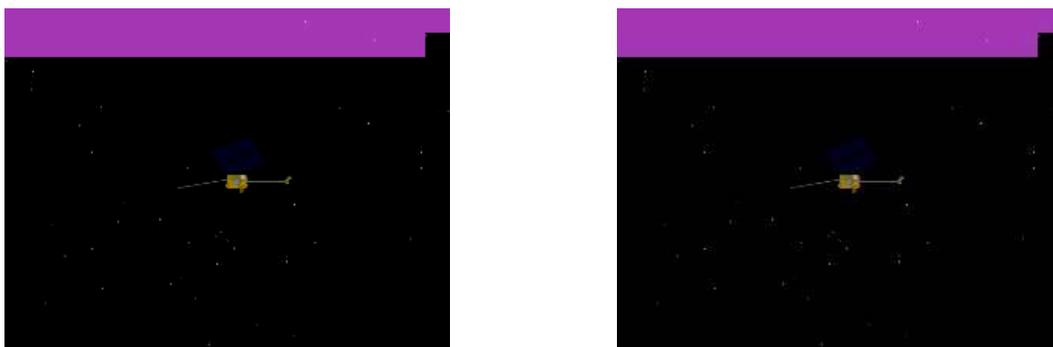
Figures 8 and 9 : left (8) : marsobserver – right (9): JPEG compressed image with quality 90%

6. APPLICATION: STIRMARK BENCHMARK

The algorithm described in this paper has become a part of the Stirmark Benchmark based on the Stirmark tool for evaluation of digital watermarking algorithms [7,8]. It is used to rate the changes of perceptual quality of images after different manipulations. As there are a number of attacks with varying parameters being applied on a watermarked image to test the robustness of the watermark, it is necessary to have some information about the quality of the resulting images: A watermark only needs to be robust against attacks not degrading the quality of an image so much that it becomes useless for the attacker. Without a computer-aided tool to decide if the perceived quality is reduced, subjective testing would be necessary. This is not possible in the online-environment of Stirmark Benchmark. Another important reason to use our algorithm is to get objective results not dependent on the viewer. Every decision is made by the same algorithm.

As stated in section 5 of this paper, the results of the algorithm ignore the fact that the perceived quality changes depend on the content of the image. In the case of image equality evaluation after attacks the results are reliable enough to be used.

7. CONCLUSION

We have introduced a computer aided visual model for checking transparency of watermarking operations on video streams. The model takes two digital images as the input of the and returns a probability that an observer can distinguish the two pictures. It is based on a visual modulation threshold function.

The different steps to implement the model in a watermarking environment have been discussed, enabling a watermark designer to use it as a test function for watermarking transparency. As the visual model has a rather large threshold area, it can not be used as a stand alone solution for quality checking. But it can be a valuable help, as many checked images give clear results. If the model can not offer an advice, this can be detected and a subjective test must take place.

Subjective tests have caused a lot of problems and discussions in recent watermarking evaluation processes as persons with varying backgrounds produce different results. This makes it hard to decide about quality. A commonly accepted software solution for this problem will make whole watermarking evaluation process much more efficient. To achieve this, the results of the algorithm have to be compared to subjective test results to prove its ability of quality evaluation.

REFERENCES

1. J. Dittmann, M. Stabenau, R. Steinmetz, *Robust MPEG Video Watermarking Technologies*, Proceedings of ACM Multimedia'98, the 6th ACM International Multimedia Conference, Bristol, England, pp. 71-80,1998
2. A. B. Watson, ed., *Digital Images and human Vision*, Masschsetts Institute of Technology, 1993
3. A. B. Watson, *Detection and Recognition of Simple Spatial Forms*, in Physical and Biological Processing of Images ed. By O. Braddick and A. Sleigh. Berlin, Springerverlag
4. P. J. Burt, E. Adelson, *The Laplacian Pyramid as a Compact Image Code*, IEEE Transactions on Communications, Vol. Com- 31, No. 4, April 1983
5. J. Dittmann, A. Mukherjee, M. Steinebach, *Media-independent Watermarking Classification and the need for combining digital video and audio watermarking for media authentication*, Proceedings of the International Conference on Information Technology: Coding and Computing, 27 - 29 March, Las Vegas, Nevada, USA, pp. 62 - 67, IEEE Computer Society, ISBN 0 - 7695 - 0540 - 6, 2000
6. J. Dittmann, A. Steinmetz, R. Steinmetz, *Content-based Digital Signature for Motion Pictures Authentication and Content-Fragile Watermarking*, Inn Proc. of IEEE Multimedia Systems, Multimedia Computing and Systems, June 7-11, 1999, Florence, Italy, Volume1, pp. 574-579, 1999
7. F. A. P. Petitcolas, R. J. Anderson, M. G. Kuhn. *Attacks on copyright marking systems*, in David Aucsmith (Ed), Information Hiding, Second International Workshop, IH'98, Portland, Oregon, U.S.A., April 15-17, 1998, Proceedings, LNCS 1525, Springer-Verlag, ISBN 3-540-65386-4, pp. 219-239
8. F. A. P. Petitcolas and R. J. Anderson, Evaluation of copyright marking systems. In proceedings of IEEE Multimedia Systems (ICMCS'99), vol. 1, pp.574--579, 7--11 June 1999, Florence, Italy
9. A.Braun, Qualitätskriterien bei digitalem Fernsehen und Video, Diplomarbeit WS 1997/98, Swiss Federal Institute of Technology Zurich
10. A. P. Ginsburg, *Spatial Filtering and Visual Form Perception*, John Wiley and Sons, New York 1986

Quality	%Size	RMS Diff	JND	Inference
10	04.80	12.8071	7.75813	Y
20	07.05	9.97017	6.47521	Y
30	09.01	8.68706	6.01971	Y
40	10.89	7.89389	5.54938	Y
50	12.85	7.27618	5.07515	Y/N
50	15.09	6.71154	4.90694	Y/N
70	18.40	5.97333	4.46516	Y/N
80	24.00	4.91071	3.99615	N
90	36.00	3.20033	3.07240	N

Table 2: Test Result for lena

Quality	% Size	RMS Diff	JND	Inference
10	12.90	9.62377	8.58123	Y
20	16.50	7.16238	6.55571	Y
30	19.30	6.12277	6.47361	Y
40	21.50	5.21551	6.0808	Y
50	23.75	4.66783	5.96942	Y
60	26.20	4.09742	5.47024	Y
70	29.50	3.48799	5.08688	Y
80	34.85	2.59926	4.86640	Y/N
90	46.11	1.42063	3.84835	N

Table 3: Test Result for dropcollision

Quality	% Size	RMS Diff	JND	Inference
10	11.40	3.12025	7.65350	Y
20	13.16	2.32563	5.90114	Y
30	14.47	1.95905	5.62897	Y
40	16.19	1.73350	4.80583	Y/N
50	17.60	1.57219	4.71805	Y/N
60	19.33	1.43228	4.32217	Y/N
70	22.17	1.24202	3.76687	N
80	26.11	1.00053	3.62986	N
90	35.73	0.70893	2.76742	N

Table 4: Test Result for canyon

Quality	% Size	RMS Diff	JND	Inference
10	05.60	7.95284	3.93110	Y
20	07.80	5.32152	4.30875	Y
30	09.80	4.25796	2.97484	Y/N
40	11.60	3.68933	2.85760	Y/N
50	13.33	3.26806	3.01054	N
60	15.20	2.91495	2.57222	N
70	18.00	2.51082	2.50305	N
80	22.20	2.06518	1	N
90	32.40	1.46336	1	N

Table 5: Test Result for antarktis

Quality	%Size	RMS Diff	JND	Inference
10	33.6	3.34353	44.4173	y
20	34.9	2.60206	38.1064	y
30	35.53	2.27424	35.2366	y
40	37.03	2.1746	42.9677	y
50	38.91	1.77787	35.4150	y
60	40.79	1.60725	36.7146	n
70	42.86	1.55047	43.3721	n
80	46.43	0.987929	36.5959	n
90	53.95	0.67977	35.7555	n

Table 6: Test Result for mar observer