MVA2005 IAPR Conference on Machine VIsion Applications, May 16-18, 2005 Tsukuba Science City, Japan
13-21

Fast quality measurement of a H263+ video stream for teleoperating a HRP-2 humanoid robot

Olivier Stasse, Neo Ee Sian and Kazuhito Yokoi
AIST/IS-CNRS/STIC Joint Japanese-French Robotics Laboratory (JRL)
Intelligent Systems Research Institute (IS),
National Institute of Advanced Industrial Science and Technology (AIST)
AIST Central 2, Umezono 1-1-1, Tsukuba, Ibaraki, 305-8568 Japan
{olivier.stasse,kazuhito.yokoi}@aist.go.jp

Gabriel Dauphin, Patrick Bonnin
Laboratoire de Transport et de Traitement de l'Information
Institut Galilée, Université Paris XIII
93430, Villetaneuse, France
dauphin@l2ti.univ-paris13.fr,bonnin@iutv.univ-paris13.fr

***Abstract*— This paper describes an experimental software platform for measuring image quality on a H263+ video stream. The goal is evaluate the lost of quality induced by the compression scheme. This architecture has been used while teleoperating a humanoid robot. Among the three quality measurements presented, a new distance is proposed to tackle the reference problem.**

***Index Terms*— H263+, image quality measurement, teleoperation**

I. Introduction

This paper presents a software platform aiming at measuring the image quality of a H263+ video stream. This platform has been realized in the context of a French national project called Cleopatre, and a collaboration between the newly funded Joint Japanese-French Robotics Laboratory and the L2TI (partner of the Cleopatre project). The software platform is a step towards the final application aiming at controlling the quality of a compressed video stream with loss of information. In the context of a teleoperation application, it can be extremely important to insure that the visual feedback is sufficiently correct for a human operator. In the Cleopatre project, the final goal is to remotely control a mobile robot.

The main difficulty, considering image quality measurement, is to find a distance coherent with the human perception. A second problem considering network applications of such measure is the reference. Indeed when receiving the compressed image it is not possible to compare it with the full initial image, otherwise this latter one would have been transmitted. In this paper the coherency with human perception is not directly addressed, indeed we rely on previous works addressing this issue. Instead a new measurement quality trying to tackle the second aspect is presented. Thus, three different distances have been considered:

1) Universal Image Quality Index [1] using the whole original image as the reference,
2) Generalised Block-Edge Impairment Metric for Video Coding [2] without any reference,
3) Edge Quality Measurement, a new metric based on edges.

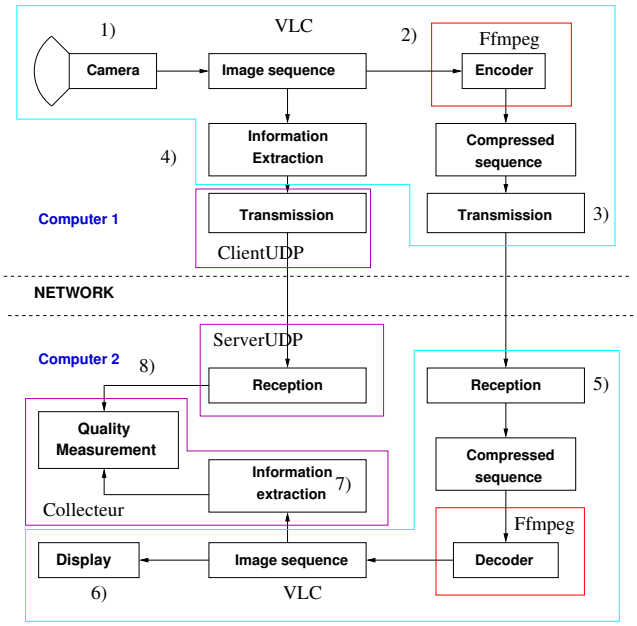

Fig. 1. Current structure of the quality measurement system

The remainder of this paper is organised as follows: Section II recalls briefly the H263+ compression scheme, and the source of quality loss. Section III describes the quality measurement used for comparisons and a novel quality measure. Section IV describes the software platform used to perform the measurements. Finally, section V reports the experimental results obtained on a humanoid robot.

II. H263+ COMPRESSION

This protocol described in [3] is based on a similar compression process than MPEG-1. It slices an image in blocks of 8x8 pixels. On each block a discrete cosinus transform is applied. Each coefficient is quantizied. Then, the coefficients for a block are sort according to the image frequency they represent. The resulting binary sequence is compressed with a lossless compression scheme similar to the Huffman algorithm. However in H263+, the dictionary is specified, and chosen to compress efficiently the coefficients which occur the most frequently. H263+ uses this scheme to send full and partial images according to the stream of images.

They are several causes for quality lost: The effect induced by image representation using the discrete cosinus transform and its 8x8 spatial support. The quantification introduces also a lost of information, and finally according to the chosen bandwidth some higher frequencies might be set to zero. One can notice that the choice of the dictionary is also relevant. Indeed if the environment contains frequencies which does not belong to the optimally compressed set, then the data packets created might induces some network congestion.

III. QUALITY MEASUREMENT

A. GBIM

The *Generalised Block-Edge Impairment Metric for Video Coding* [2] measures mainly the deformation induces by the block partionning. This measure does not need any reference, and uses only the gray level. It is based upon the gradient applied to the blocks' edges. The proposed distance takes into account some particularities of the human visual system. In the following the gray level intensity of the compressed image at point (i, j) is noted $g^C_{i,j}$, while it is noted $g^S_{i,j}$ for the initiale image, also called "source". In the remainder the size of an image is $m \times n$.

1) Building the weighting function $w_{i,k}$: This weighting function is a matrix of size $n \times m/(8-1)$. It is build upon the functions $\mu_{i,k}$ and $\sigma_{i,k}$. $\mu_{i,k}$ is the average on the 8-left pixels, and the 8-right pixels:

$$\mu_{i,k} = \frac{\mu^L_{i,k} + \mu^L_{i,k}}{2}, \tag{1}$$

with

$$\begin{aligned} \mu^L_{i,k} &= \frac{1}{8} \sum_{j=8\times k-7}^{8\times k} g^C_{i,j} \\ \mu^L_{i,k} &= \frac{1}{8} \sum_{j=8\times k+1}^{8\times k+8} g^C_{i,j} \end{aligned} \tag{2}$$

In same manner, $\sigma_{i,k}$ is computed from the 8 left and right pixels:

$$\sigma_{i,k} = \frac{\sigma^L_{i,k} + \sigma^R_{i,k}}{2} \tag{3}$$

with

$$\begin{aligned} \sigma^L_{i,k} &= \left[1/8 \sum_{j=8\times k-7}^{8\times k} (g^C_{i,j} - \mu^L_{i,k})^2 \right]^{0.5} \\ \sigma^R_{i,k} &= \left[1/8 \sum_{j=8\times k-7}^{8\times k} (g^C_{i,j} - \mu^R_{i,k})^2 \right]^{0.5} \end{aligned} \tag{4}$$

Finally the weighting function is build as:

$$\begin{aligned} w_{i,k} &= \lambda ln(1 + \frac{\sqrt(\mu_{i,k})}{1+\sigma_{i,k}}) \quad \text{if } \mu_{i,k} \leq \zeta \\ w_{i,k} &= ln(1 + \frac{\sqrt(\mu_{i,k})}{1+\sigma_{i,k}}) \quad \text{otherwise} \end{aligned} \tag{5}$$

where $\zeta = 81/255$ and $\lambda = \frac{ln(1+\sqrt{1-\zeta})}{ln(1+\sqrt{\zeta}}$)

2) Computing the block effect from the variation between columns: The block effect in the horizontal direction is the sum of the squared differences between borders columns.

$$Mh^2 = \frac{1}{m(n/8-1)} \sum_{i=1}^{m} \sum_{k=1}^{n/8-1} w_{i,k} (g^C_{i,8\times k} - g^C_{i,8\times k+1})^2 \tag{6}$$

This quantity is then normalised in order to take into account the variations in others columns:

$$MhGBIM = \frac{Mh}{S_n} \text{ where } E = \frac{1}{7}(\sum_{n=1}^{7} S_n) \tag{7}$$

and

$$S^2_n = \frac{1}{m(n/8-1)} \sum_{i=1}^{m} \sum_{k=1}^{n/8-1} w_{i,k} (g^C_{i,8\times k+n} - g^C_{i,8\times k+n+1})^2 \tag{8}$$

3) Measuring the block effect from the variation between lines: The block effect measurement in the vertical direction is done by applying the previous steps (paragraphs III-A.1 and III-A.2) to the transposed image: $g'_{i,j} = g_{j,i}$. This provides *MvGBIM* instead of *MhGBIM*. The final measurement is:

$$MGBIM = \frac{MvGBIM + MhGBIM}{2} \tag{9}$$

As this measurement takes its values on $[1, \infty[$, the following computation provides a quality measurement on $[0, 1]$:

$$GBIM = \frac{1}{MGBIM} \tag{10}$$

B. UIQ

In the other hand, the *Universal Image Quality Index* [1] does not take into account such kind of information. This measure is statistically grounded and uses the mean and the variance between pixels of the distorted image and the initial one.

The two images are splitted in blocks of 8 by 8 pixels. Correlation and distortion for both luminance and contrast are computed on each block:

$$Q_{bloc} = \frac{\sigma_{g^S_{i,j} g^C_{i,j}}}{\sigma_{g^S_{i,j}} \sigma_{g^C_{i,j}}} \times \frac{g^S_{i,j} g^C_{i,j}}{(g^S_{i,j})^2 + (g^C_{i,j})^2} \times \frac{2}{\sigma^2_{g^S_{i,j}} + \sigma^2_{g^S_{i,j}}} \tag{11}$$

where $g^S_{i,j}$ is the gray level for each pixel of a block in the source image, and $g^C_{i,j}$ the corresponding gray level in the compressed image. $g^S_{i,j}$ and $g^C_{i,j}$ represent the average of x and y, $\sigma_{g^S_{i,j}}$ and $\sigma_{g^C_{i,j}}$ are the variances for $g^S_{i,j}$ and $g^C_{i,j}$.

Each value Q_{bloc} is then averaged on the overall image. The result is given between $[-1;1]$ with 1 corresponding to an image without loss. An affine transformation put those values between 0 and 1. The measure obtained is noted *UIQ*:

$$UIQ = \frac{1}{2}(1 + \frac{8\times 8}{n\times m}\sum_{block} Q_{block})$$

Initially designed for a gray-level image, it is applied to the 3 channels : YUV.

C. EQM

The quality measurement proposed here computes a distance between the edges of the initial image and the reconstructed one. The edges are computed on the 3 color channels using a Kirsh operator. A "valley image" is generated by using a paraboloid on the binary image $(b_{i,j})$ representing the edges. The paraboloid is defined as $p_{i,j} = (i^2 + j^2)$ for $|i| \leq t$ and $|j| \leq t$, 0 otherwise. Then the valley image is the limit of the images of size $m \times n$ with $v^0_{i,j} = t^2$ and

$$\forall(k,l) \text{ such as } b_{k,l} = 1, v^{k+1}_{i,j} = min(v^k_{i,j}, p_{k-i,l-j}) \quad (12)$$

The limit of this serie of images is noted

$$d_{i,j} = v^\infty_{i,j} \quad (13)$$

and form a set of distances. The global distance is then computed with

$$EQM = \sum_{(i,j)|b_{i,j}=1} \frac{1}{1+\alpha d_{i,j}} \quad (14)$$

IV. The software platform

A. Overview

The general scheme of this experience depicted in figure 1 is the following: 1) A sequence of images grabbed from the camera, is send to a coder. 2) This coder compressed the sequence of images using an algorithm with information loss. 3) The subsequent data are send over the network to a remote computer. 4) The sequence of images is processed to extract supplementary information to compute the quality measurement. 5) On the remote computer the compressed data are used to reconstruct partially the original images. 6) The subsequent images are displayed. 7) The same process used in the video stream is applied to the obtained images. 8) The result of both processes are used to compute a distance.

Most of the video streaming parts is handled by VideoLan Client (VLC) [4] an Open Source project aiming at sending and receiving multimedia streams such as MPEG-2, MPEG-4, satellites channels, and so on. This program has been modified to fit our needs. For H263+ the encoding is realized by FFMPEG. This Open Source project integrates several protocols for video transmission.

Two programs ClientUDP and ServerUDP take care of transmitting the information extracted from the initial image sequence, and from the reconstructed image sequence. A third program called the Collecteur synchronise the arrival of the informations, and computes the image quality.

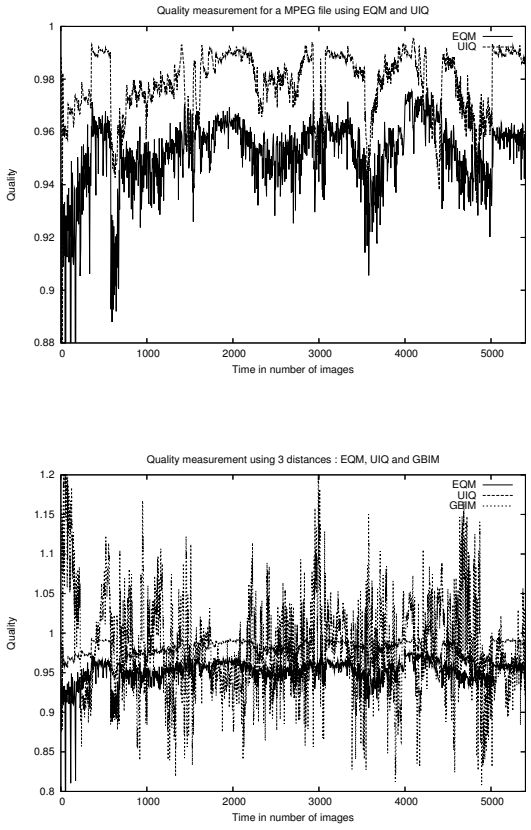

Fig. 2. Quality measurement of a MPEG file for UIQ, EQM (up) and GBIM (bottom)

V. Experiments

A. Simple experiment for comparison

This software platform has been installed between two computers on a local area network. In order to be able to compare the three measurements algorithms the same MPEG file has been played for each of them. The evolution of UIQ and EQM is reported in figure 2, while the 3 measures are depicted together in figure 2. This preliminary result indicates that the newly introduced measurement (EQM) using edges is similar to the one using the whole image (UIQ). UIQ and EQM have a correlation of 0.0000697 while UIQ and GBIM have a correlation of -0.00265. Moreover while UIQ takes 100 milli seconds to compute, EQM takes 10 milli seconds. EQM is thus suitable to deal with a stream of images coming from a robot, and gives similar result on this sample than UIQ.

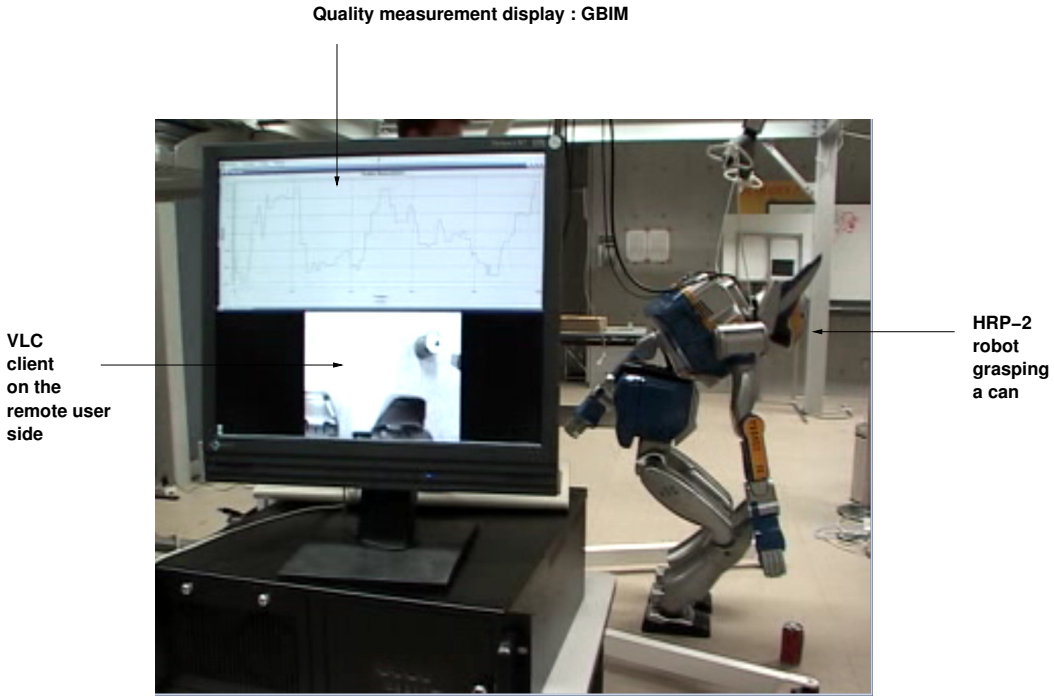

Fig. 3. Teleoperation experience using the HRP-2 humanoid robot

B. Experiment using a humanoid robot

This software platform has also been modified to be used together with the HRP-2 humanoid robot [5]. This robot has 32 degrees of freedom, weight 54 kilo grams, and has a trinocular camera system. HRP-2 is able to achieve a complex sequence of actions such as searching for a can, locating it accurately, grasping it, walking towards a garbage and trash it by a mixture of teleoperation [6] and autonomous functions [7]. Using a full body motion framework [6] it is possible to control the end-effectors simply through joysticks. The main feedback for the remote user is through direct eye-contact and video stream.

In this experiment both have been used. The VLC server is running on the PC dedicated to the vision side, whereas the teleoperation framework part of the robot works on the PC dedicated to control. As a first step the robot is connected using a 100 Mbit/s connection with the remote user. In this configuration, the three quality measurement have been working without any congestion problem.

In a second step the robot has been connected using a 802.11b 10 Mbit/s connection. In this configuration, unfortunately neither UIQ neither EQM could be used because of the network traffic involved with the teleoperation system. Only GBIM as it is reference free, was running without any problem. The wireless setup is depicted in figure 3. Another major problem related to this current implementation is the delay between the acquisition time and the rendering on the display. The compression-decompression-buffering-rendering sequence takes 2 seconds. Also that is sufficient for giving high-level abstract order, this is not good enough to improve the skills of the robot.

VI. Conclusion

In this paper a novel and fast quality measurement has been proposed. It allows frame-rate computation and offer results similar to UIQ, a measurement statistically grounded and based on the original image. This quality measurement has been implemented and tested on a H263+ video stream while teleoperating a humanoid robot.

Our future work will consist in decreasing the delay between the acquisition time, and the rendering. Moreover the possibility of controlling the compression implementation to stabilise the image quality might help the user in real applications.

References

[1] Z.Wang and A.C.Bovik, "A universal image quality index," *IEEE Signal Processing Letters*, vol. XX, March 2002.

[2] H.R.Wu and M.Yuen, "A generalized block-edge impairment metric for video coding," *IEEE Signal Processing Letters*, vol. 4, november 1997.

[3] R. ITU-T, *H.263, 1998, H.263 Series H: AUDIOVISUAL AND MULTIMEDIA SYSTEMS, Infrastructure of audiovisual services - Coding of moving video*, 02 1998.

[4] E. C. de Paris, *The VideoLan Project*. http://www.videolan.org/, 2003.

[5] K.Kaneko, F.Kanehiro, S.Kajita, H.Hirukawa, T.Kawasaki, M.Hirata, K.Akachi, and T.Isozumi, "Humanoid robot hrp-2," in *Proceedings of the 2004 IEEE International Conference on Robotics & Automation*, 2004.

[6] N. E. Sian, K. Yokoi, S. Kajita, and K. Tanie, "A framework for remote execution of whole body motions for humanoid robots," in *International Conference on Humanoid Robots, Los Angeles*, November 2004.

[7] Y. Kawai, Y. Fukase, F. Tomita, and H. Hirukawa, "A stereo vision system for the hrp-2 humanoid robot to act autonomously," in *Asian Conference on Computer Vision*, vol. 2, 2004, pp. 754–759.