CHAPITRE II-2

PREMIERE ETUDE D’EVALUATION DIAGNOSTIQUE
Chapitre II-2: Première étude d’évaluation diagnostique

Préliminaires

L’application de la compression avec perte dans le domaine médical requiert une évaluation sur le contenu diagnostic des images comprimées. Nous avons conduit une première campagne d’évaluation basée sur les interprétations diagnostiques d’angiographies complètes avec ou sans compression.

Ces travaux ont fait l’objet d’une rédaction en anglais en vue de leur communication au sein de la Société Philips, et en vue d’une publication qui sera soumis dans les prochains mois dans une revue cardiologique.


1. INTRODUCTION

The assessment of images of the heart or of the coronary arteries for diagnostic purposes or to determine the feasibility of therapeutic interventions is usually based on cine-film recordings. Even when the catheterization laboratory images were originally digitized, they are still not stored in a digital format due to technological and financial constraints: a standard cardiac angiogram comprises 1000 to 3000 images, representing 250 to 750 Mega Bytes for 512x512 pixel images coded in 8 bits. Image data compression is necessary to achieve fully digitized archiving of cardiac angiography and to enable images to be transmitted via networks.

Real time lossless data compression has already been available for several years on the Philips Digital Cardiac Imaging (DCI) system. It doubles the disk capacity of the angiography system. But this additional capacity is still not sufficient to store the daily work load of a catheterization laboratory. Lossfree compression is generally associated with reduction factors or compression ratios (CR) of 2:1 for this type of image [ROOS-88]. This means that two compressed images can be stored on the same disk space as one single non compressed image. For full storage of patient image data, we estimated that a compression ratio of 8 to 16:1 was necessary. Such CRs are only possible with lossy data compression methods, implying that the image is irreversibly changed. One must therefore ensure in clinical conditions that the diagnostic quality of cardiac angiograms is not altered after lossy data compression.

We performed a study on 18 patients to evaluate whether cardiac angiograms after lossy compression could replace the originals in the working catheterization laboratory environment. Two compression algorithms were assessed with a reduction factor of 12:1: the
standard MPEG technique, and a scheme based on the so-called MLOT transform specially developed for this study. Two observers reviewed separately the digital films. Each patient angiogram was displayed at least three times in the course of the study, in its original, MPEG compressed, and MLOT compressed version. The interpretation was performed using the standard procedures of the Service d'Hémodynamique in the Hôpital Cardiologique, Lille, France.

2. MATERIALS & METHODS

2.1. Patient selection

The full digital cardiac angiograms of 18 patients were acquired from a mono-plane DCI system. Each angiogram included a ventriculogram and a coronary angiogram. Only diagnostic procedures were studied. Purely interventional procedures were not included. If a diagnostic procedure was followed by an intervention, the intervention images were removed. Patients with normal coronary arteries, with single, double, or triple vessel disease were studied.

2.2. Image selection

In Lille, a diagnostic examination follows a standardized format. The left ventriculogram is obtained in the 9° Image Intensifier (II) mode with a Right Anterior Oblique 30° projection (RAO-30); sometimes a Left Anterior Oblique 45° (LAO-45) is also performed. The left coronary artery is filmed in the 7° II mode in six projections: RAO-30, RAO-15/Ca15 (Ca means caudal), RAO-10/Cr15 (Cr means cranial), RAO-10/Cr40, LAO-55/Cr25, and LAO 90. The right coronary is viewed in the 7° II mode in three projections: LAO-90, LAO-45, RAO-45. Sometimes artery segments are magnified in 5° II mode or filmed in other projections in order to further assess a lesion. The film speed is always 12.5 images per second for both the LV and coronary imaging. We removed some images of no diagnostic use such as those with no contrast medium, resulting in an average number of images per patient of 580, ranging from 430 to 740; this is rather low compared to many catheterization laboratories where the acquisition speed is higher. Images were exported from a DCI system into a Sun platform with the PMSnet™ networking. They were converted from their ACR-NEMA format to Unix files and stored on digital tapes.

2.3. Image processing

All images from the 18 patient angiograms were compressed and reconstructed with the MPEG and MLOT algorithms, at a compression ratio of 12:1. We will later refer to MPEG$_{12}$ or MLOT$_{12}$ compressed images or angiograms. In this paper the word "compressed" refers to an image after it has been compressed and reconstructed. Strictly speaking, the compressed image is a binary file of small size that cannot be visualized; the reconstructed image is a pixel file with the same size as the original and which can be visualized. The words coded/coding are here the same meaning as compressed/compression.

Block-based transform coding is currently the most widespread lossy compression technique. Some standards are now commonly used. The Moving Picture Experts Group (MPEG) has proposed a video compression standard for multimedia applications [LEGA-91]. Briefly, the
redundancy from one frame to the other is reduced by a block-based motion compensation; the spatial or intra-frame redundancy is reduced by quantizing the frequency components derived from a particular mathematical transform of the image pixels: the block Discrete Cosine Transform (DCT). A disadvantage of block-based compression lies in the fact that individual blocks become visible at high compression ratios. This is particularly true for medical images and cardiac angiograms because of the viewing conditions. The latter tends to enhance the contours of the vessels, and in the same time the block boundaries. To overcome the so-called "blocking effect" or "block artifact", a dedicated algorithm was investigated. It is an intra-frame coding based on the Modified Lapped Orthogonal Transform (MLOT) [BREE-94]. In the process of transforming the pixels from spatial gray values into frequency coefficients, some information from adjacent blocks is used. This does not generate blocking in the reconstructed image. In the development of the MLOT algorithm, the quantization of frequency coefficients has been adapted to the viewing conditions of the catheterization laboratory, and especially to the edge enhancement applied by default in the digital cardiac angiography DCI system. This edge enhancement is a convolution filter. It increases the medium and high frequencies of the image. This makes the vessel contours and the small vessels more visible, at the cost of increasing the visibility of the image noise and ... enhancing some compression artifacts.

2.4. Observers and viewing sessions

Non-compressed and compressed angiograms were interpreted by two observers: two experienced angiographers and interventionalists, one junior and one senior. The viewing sessions took place in the normal clinical environment. A viewing system was built and installed there for the study; it has exactly the same operator console and display devices as the catheterization laboratory DCIs. Typically, a viewing session took 45 minutes, the time needed to review, interpret and report on 9 patient angiograms. Of the 9 sets of patient images reviewed in a single session, some were original, some were MPEG12 and some were MLOT12; the observer was not informed which was which. In the course of the study, both observers interpreted separately the 18 original angiograms, the 18 MPEG12 and the 18 MLOT12 ones. In addition, 5 original, 5 MPEG12, and 5 MLOT12 angiograms were also reviewed a second time to analyse the intra-observer variability. Ideally we should have repeated all 18 originals, 18 MPEG12 and 18 MLOT12 angiograms but this was impossible for time constraints. About 35 000 images were handled for this study. Finally, 69 angiograms were interpreted by the two observers, resulting in a total of 138 interpretations. Observers were free to review each film at their own pace, to move forward and backward and change the speed. But changing the post-processing settings was not allowed, such as applying zoom, increasing the default edge enhancement, changing contrast and brightness. All these have an influence on the visibility of compression artifacts, especially for MPEG, and it would be a non-reproducible factor making the analysis of our data very difficult.

2.5. Diagnostic task

In interpreting the angiograms of this study, observers were asked to do the same diagnostic task as in the clinical practice at their Cardiology Department. A sheet was derived from the information recorded after every catheterization laboratory procedure. This sheet or diagnostic questionnaire included the interpretation of the left ventricle and the coronary arteries (see appendix). The cardiologists had to locate and grade the abnormalities of the LV and the coronary artery segments. The segments were classified using the CASS system.
currently used in the Department. No patient information was provided, the diagnostic questionnaires were filled in using the image information only. Briefly, the diagnostic tasks performed on the angiograms by cardiologists consists in assessing the Left Ventricle (LV) function, and the presence and severity of coronary disease. The ventriculogram shows the dynamics of the LV. If an LV segment does not contract properly, the ability of the heart to eject blood to the body is reduced. Doctors judge the contraction patterns of the LV walls. Absence of contraction is in many cases due to coronary disease. The coronary arteriogram shows the coronary arteries lumen. Doctors look for narrowing, partial or total occlusions of the main coronary branches.

3. Statistical analysis

3.1. Introduction

Angiography is generally considered as the definitive procedure for coronary disease. The only standard available for assessing the diagnostic value of a compressed angiogram is the non-compressed angiogram. We compare a new imaging technique with an established one to see whether they agree sufficiently for the new to replace the former. Comparing the new technique with the "true" diagnostic is not possible. The main objective for the statistical analysis of the diagnostic questionnaires was to assess the degree of agreement between the compressed and the original angiograms. The protocol of this study is novator in many respects: review of digital angiogram on a catheterization laboratory system, not a computer; assessment of the diagnostic quality of full cardiac angiograms; use of the ordinary diagnostic tasks. We had to design a new approach in order to assess the impact of lossy data compression on the cardiologists' diagnostic from cardiac angiograms. An exam consists of diagnostic tasks, for the LV and the coronary artery. We have analyzed them separately.

3.2. Analysis of the LV interpretations

The LV segments assessed for the diagnostic questionnaire were the anterior, apical, inferior and lateral ones. For each segment, the observers had to classify wall motion as: "normal", "akinetische", "hypokinetic", "dyskinetic". This type of rating is categorical. An appropriate measurement of agreement for this type of data is the kappa introduced by [COHE-60] and described by [FLEI-81]. Kappa compares the degree of agreement between two ratings with the agreement from chance alone; it sample estimate is denoted here $k^\wedge$. If there is complete agreement $k^\wedge=1$; if the observed agreement is greater than chance $k^\wedge>0$; if it is less than chance $k^\wedge<0$; the minimum value of $k^\wedge$ lies between -1 and 0.

We computed the kappa estimates for the following agreements between interpretations of angiograms: interpretations on original compared with interpretations on compressed films for one observer; interpretation by observer 1 of the original and by observer 2 of the compressed films; interpretation by observer 1 and by observer 2 of the same set of films (original or compressed). We tested whether the kappa estimates were statistically non equal to zero i.e. whether an observed agreement is greater than chance. We looked if the agreements on original interpretation were significantly different than the other agreements. The evaluation of the kappa estimates and the tests on their significance and on comparison of several kappas are given in [FLEIS-81].
3.3. Analysis of the coronary angiogram interpretations

On each diagnostic questionnaire, the observers graded 15 coronary segments and 4 possible grafts. The vessel diameter narrowing were graded from 0 (normal) to 5 (occlusion), with 3 sub-divisions for occlusion using the TIMI system. A correspondence of the grades and percent stenosis is given in table 7, section 3-3-1

Like for the ventriculograms, we had to estimate the agreements between the interpretations of original and compressed films, or between the interpretations of observer 1 and observer 2. The rating is here ordinal and will be treated like a numerical value. As explained in section 3-3-1, we had to modify the grades into global scores for the main coronary branches. Nevertheless, the principle of the analysis remained unchanged.

For comparing interpretations, the first step was to plot the data of one set of interpretations versus the data of the second set. If the interpretation of each patient were all identical, all the points would fall on the line of equality. A simple observation of the data on such a plot is very informative.

The second step is to summarize the agreement gauged by the eye into an appropriate statistics. A natural approach seems to be the estimate of the Pearson correlation coefficient but it is misleading here [BLAN-86], [LIN-89]. The Pearson correlation coefficient measures the strength of a linear relationship between two sets of grading, not the agreement between them. Another widespread method is the paired t-test. It evaluates whether the estimated means of the two sets of grading are equal, this may be true even with a very poor agreement.

We chose a concordance correlation coefficient proposed by Lin [LIN-89]. It evaluates the degree to which pairs of grading fall on the line of equality. It is scaled between -1 and +1. +1 is the perfect agreement, -1 the perfect reversed agreement, and 0 the absence of agreement (i.e. the grading are independent). We computed the estimates and derived the confidence intervals from the estimates of the variance under a gaussian assumption.

4. RESULTS

4.1. Image quality judgements

Observers reviewed original and compressed angiograms. They were aware of this fact and had been informed of the purpose of the study. But to their surprise they were never able to tell whether a film was compressed or not, they could see nothing in the images that would make compression distinguishable, neither for MPEG, nor for MLOT. This is a very important qualitative result: in the default viewing conditions of the catheterization laboratory, MPEG or MLOT compression at a compression ratio of 12:1 were not detected by the doctors.

A side item was asked in the diagnostic questionnaire: observers had to give their global opinion of the image quality on a three-point scale (good, average, poor). The grade "poor" was never used for any of the 138 diagnostic questionnaires, and the grade "good" was the most common. On 46 interpretations of the original angiograms, 7 were "average"; on 46 interpretations of the MPEG12 angiograms, 12 were "average"; and on the 46 interpretations of the MLOT12, 8 were "average". The MPEG12 films seem to be slightly worse than the original and MLOT12 ones, but this differences are not statistically significant ($\chi^2$ test :
The observed numbers of "average" on original, MPEG\textsubscript{12} or MLOT\textsubscript{12} are not statistically different.

The opinion of the investigators, who have a technical background, provides an additional qualitative result. To their eyes the MPEG\textsubscript{12} films could be recognized because some blocking structures were clearly visible in the background of the images. This was probably not noted by the doctors as there is no diagnostic information in the background. During dynamic viewing it gave the impression of moving square patterns, and a horizontal and vertical texture could be seen on still frames. As the actual frequency contents of each block is different and as blocks are coded separately, one can see that some blocks are more blurred than others. Moreover, a slight haziness can be seen on MPEG\textsubscript{12} images for tiny vessels. It is harder to see in large vessels in the default viewing conditions. One can also detect that a film is MLOT\textsubscript{12} coded but this requires much closer attention. Only a very slight difference in the usual noise pattern or noise texture can be seen. If the default viewing conditions are changed, by applying zoom for instance, MPEG\textsubscript{12} images look poor: very "grainy" (blocking effect) and blurred, the contours of large vessels are not very well defined and would be difficult to draw in a quantization software. Zoomed MLOT\textsubscript{12} images were of good quality, slightly lacking in definition.

4.2. Left ventricle interpretations

4.2.1. Introduction

The two cardiologists assessed four LV segments (anterior, apical, inferior, lateral). The lateral segment is only seen in the LAO-45 projection which was performed in only 4 of the 18 patient angiograms. Thus data for the lateral segments was not analyzed. The category "dyskinetic" was never used by the observer and will not appear in the analysis. Classification of abnormalities on medical images is subject to inter and intra-observer variability. We focused on assessing whether the possible variability caused by lossy data compression in the ventriculogram interpretations was comparable to the observer variability.

4.2.2. Inter-observer agreement

We firstly analyzed the inter-observers agreement resulting from the interpretations of the original angiograms by the two observers, as shown in table 1. We indicated “Npatient”, the number of patients interpreted for each LV segment; “Ndisag”, the number of disagreements; the kappa estimate, and “p”, the p-value from the test of independence.
Chapitre II-2: Première étude d’évaluation diagnostique

<table>
<thead>
<tr>
<th>LV segment</th>
<th>Npatient</th>
<th>Ndisag</th>
<th>k,^</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anterior</td>
<td>18</td>
<td>2</td>
<td>0.73</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Apical</td>
<td>18</td>
<td>4</td>
<td>0.65</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Inferior</td>
<td>18</td>
<td>5</td>
<td>0.52</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

Table 1 : LV Inter-observers agreements on the interpretations of the original angiograms

The p-values of table 1 show that the kappa estimates for the three segments are statistically significant: observers agree more than chance alone. The agreement is good for the anterior and apical segments (k,^ >0.6); it is average on the inferior segment (0.4<k,^ <0.6).

Table 2, and 3 show the inter-observer agreement when one observer reviewed the originals and the second observer reviewed compressed films. Table 4 shows inter-observer agreements on compressed films when both observers reviewed the MPEG12 or the MLOT12 films.

<table>
<thead>
<tr>
<th>LV segment</th>
<th>Npatient</th>
<th>Ndisag</th>
<th>k,^</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anterior</td>
<td>18</td>
<td>3</td>
<td>0.56</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Apical</td>
<td>18</td>
<td>4</td>
<td>0.64</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Inferior</td>
<td>18</td>
<td>8</td>
<td>0.22</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

Table 2 : Inter-observer agreements on the interpretations of the original and the MPEG12 angiograms

<table>
<thead>
<tr>
<th>LV segment</th>
<th>Npatient</th>
<th>Ndisag</th>
<th>k,^</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anterior</td>
<td>18</td>
<td>3</td>
<td>0.56</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Apical</td>
<td>18</td>
<td>5</td>
<td>0.57</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Inferior</td>
<td>18</td>
<td>6</td>
<td>0.40</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

Table 3 : Inter-observer agreements on the interpretations of the original and the MLOT12 angiograms

<table>
<thead>
<tr>
<th>LV segment</th>
<th>Npatient</th>
<th>Ndisag</th>
<th>k,^</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anterior</td>
<td>18</td>
<td>3</td>
<td>0.56</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Apical</td>
<td>18</td>
<td>5</td>
<td>0.64</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Inferior</td>
<td>18</td>
<td>6</td>
<td>0.35</td>
<td>0.089</td>
</tr>
</tbody>
</table>

Table 4 : Inter-observer agreements on the interpretations of the compressed angiograms by both observers

The estimated kappa values are statistically significant, except for the inferior wall with originals for observer 1 and MPEG12 for observer 2.
Chapitre II-2: Première étude d’évaluation diagnostique

The kappa estimates are statistically significant (the agreement is beyond chance alone), except on the inferior segment when observer 1 interprets the originals and observer 2 the MPEG\(_{12}\) films \((p=0.089)\). The kappa estimates seem globally smaller when one of the two interpretations was done on compressed films than when they were both on originals. For each LV segment, we tested whether the estimated kappa values from original and compressed interpretations were statistically significant. The results from these tests is given hereafter for the following sets of kappas:
- \(k,^\wedge\) from originals by both observers, \(k,^\wedge\) from interpretations on originals by observer 1 and on MPEG\(_{12}\) by observer 2, \(k,^\wedge\) from interpretations on MPEG\(_{12}\) by observer 1 and original by observer 2;
the three underlying values of kappa are not statistically different \((p>0.25)\);
- \(k,^\wedge\) from originals by both observers, \(k,^\wedge\) from interpretations on originals by observer 1 and on MLOT\(_{12}\) by observer 2, \(k,^\wedge\) from interpretations on MLOT\(_{12}\) by observer 1 and original by observer 2;
the three underlying values of kappa are not statistically different \((p>0.25)\);
- \(k,^\wedge\) from originals by both observers, \(k,^\wedge\) from MPEG\(_{12}\) by both observers, \(k,^\wedge\) from MLOT\(_{12}\) by both observers;
the three underlying values of kappa are not statistically different \((p>0.25)\)

There is no statistically significant difference in the inter-observer agreement from the original and the MPEG\(_{12}\) films, or from the original and the MLOT\(_{12}\) films, or from the original, MPEG\(_{12}\), and MLOT\(_{12}\) films.

In conclusion, the agreement between the two observers on the LV segments is globally good or average, and beyond chance alone. No statistically significant difference in the inter-observer agreement was found, whether the observers both interpreted original angiograms, whether one of the two observers interpreted compressed angiograms (MPEG\(_{12}\) or MLOT\(_{12}\)), or whether the two observers interpreted compressed angiograms (MPEG\(_{12}\) or MLOT\(_{12}\)).

### 4.2.3. Intra-observer agreement

To complete the analysis of the LV diagnostic questionnaire, the intra-observer agreement could be assessed. Unfortunately the five repeated interpretations realized on the original angiogram is not sufficient to evaluate the intra-observer agreement on original angiograms (the problem is identical for the intra-observer agreement on the MPEG\(_{12}\) and the MLOT\(_{12}\) angiograms).

We have estimated the intra-observer agreement between the interpretation on originals and the interpretation on compressed angiogram, as shown in table 6.
Chapitre II-2: Première étude d’évaluation diagnostique

Table 5 : Intra-observer agreements on the interpretations of the original versus the compressed angiograms for observer 1.

<table>
<thead>
<tr>
<th>LV segment</th>
<th>Npatient</th>
<th>Ndisag</th>
<th>k,^p</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anterior</td>
<td>18</td>
<td>2</td>
<td>0.75</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Apical</td>
<td>18</td>
<td>5</td>
<td>0.54</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Inferior</td>
<td>18</td>
<td>1</td>
<td>0.91</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

Observer 1: original / observer 1: MPEG_{12}

<table>
<thead>
<tr>
<th>LV segment</th>
<th>Npatient</th>
<th>Ndisag</th>
<th>k,^p</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anterior</td>
<td>18</td>
<td>2</td>
<td>0.74</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Apical</td>
<td>18</td>
<td>3</td>
<td>0.71</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Inferior</td>
<td>18</td>
<td>3</td>
<td>0.72</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

Observer 1: original / observer 1: MLOT_{12}

A test comparing the four underlying kappa values of these intra-observer agreements showed that they were not statistically different for the anterior and apical segment (p-value>0.25), and for the inferior segment (p-value~0.20).

4.3. Coronary arteriogram interpretations

4.3.1. A score for coronary branches

The two observers assessed 15 coronary artery segments and 4 grafts in each diagnostic questionnaire. As only 1 patient had by-pass surgery, the graft segments were not analyzed. 18 segments times 15 patients had to be compared. Unfortunately, some inconstitency in the lesion location was observed, because of the observer variability and some ambiguities of the CASS system for normal anatomical variants. This can be very misleading in our data analysis. For example, if a grade 4 lesion was reported on the proximal Left Anterior Descending Artery (LAD) on the original film, and on the mid LAD on the compressed film, one would conclude on a severe disagreement: grade 4 compared to grade 0 on the proximal LAD, and grade 0 compared to grade 4 on the mid LAD. However, the two diagnostic questionnaires are relatively equivalent. To overcome this problem, a method for comparing the coronary artery interpretations of diagnostic questionnaires was needed in a way that reflects the medical conclusions. We grouped together the segments from the same functional coronary branch: four segments for the Right Coronary Artery (RCA), six segments for the LAD, and five segment for the Circumflex (CX) artery. We needed to summarize the 4 to 6 grades of each branch by a score of diagnostic use. We chose to look at the maximum grade from the 4 to 6 segments of a branch. Table 7 gives the correspondences between the scores, the grades and the diameter percentage of stenosis.
TABLE 7 : Coronary branch scores and segments grades.

Other scores based on the number of diseased segments or on a ponderation of the segments grades were tested but were of lesser diagnostic meaning. By grouping 15 segments into 3 branches, we reduced the statistical power of the analysis. This is the price we paid for experimenting this new protocol and staying as close as possible to the clinical practice. We also switched from a scale of 8 grades to a scale of 4 scores. The objective was to focus on the major major steps of the diagnostic: first detect a lesion and then classify it as significant or not. The therapeutic incidence of this decision is high whereas a more refined classification of the lesion in grade 3 or 4 for example has little consequences. The analysis reported here is based on the 4-level score we designed. A sharper analysis could also be conducted by using the maximum segment grade of each branch instead of the score we designed.

Like for the LV diagnostic questionnaires, we tried to assess whether the potential variability introduced by lossy data compression on the interpretation of coronary arteriograms is significantly higher than the observers' variability.

4.3.2. Inter-observer agreement

The inter-observer agreement on the interpretation of original coronary angiograms is an important information. It can be observed visually for each branch on a MLOT: the scores given by observer 1 on the x axis, and the scores given by observer 2 on the y axis. Ideally, all points should fall on the line of equality. The degree to which pairs of interpretation fall on this line is evaluated by the estimate of the Lin concordance correlation coefficient \( r_c \). For the three coronary branches, the inter-observer agreements on originals and their confidence interval are shown in table 8. We also indicate “\( N_{Mdisag} \)”, the number of so-called major disagreements (score 0 against score 1 is not a major disagreement), and “\( N_{Tdisag} \)”, the total number of disagreements (including score 0 against score 1).
Chapitre II-2: Première étude d’évaluation diagnostique

Table 8: Inter-observer agreement on the interpretation of original coronary arteriograms.

The agreement between the two observers in their interpretation of original arteriograms is good. The agreement of the three coronary branches are not statistically different.

Table 9 and 10 summarize the inter-observer agreements when one observer interprets the original angiograms and the second observer interprets the compressed angiogram.

Table 9: Inter-observer agreement for originals and MPEG12

Table 10: Inter-observer agreement for originals and MLOT12

These tables show that the total number of major disagreements between the interpretations by the two observers range from 3 to 7 out 18 pairs of comparisons. The estimates of the concordance correlation coefficients $r_c$ are all in the same range, and no statistically significant difference can be observed.

In conclusion, lossy compression did not change statistically the inter-observer agreements when one observer interpreted the original and the other interpreted the compressed angiograms.

4.3.3. Intra-observer agreement

The assessment of intra-observer agreement is a good complement to the preceding conclusion. We were not able to measure the intra-observer agreements on the original angiograms because the number of repeated interpretation was too small. Table 11 and 12 show the intra-observer agreement when one observer interpreted the original films and the second observer interpreted the compressed films.
### Table 11: Coronary intra-observer agreements for observer 1.

<table>
<thead>
<tr>
<th>Observer 1: original / observer 2: MPEG\textsubscript{12}</th>
<th>Observer 1: original / observer 2: MLOT\textsubscript{12}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coro. Branch</td>
<td>N\textsubscript{patient}</td>
</tr>
<tr>
<td>RCA</td>
<td>18</td>
</tr>
<tr>
<td>LAD</td>
<td>18</td>
</tr>
<tr>
<td>CX</td>
<td>18</td>
</tr>
</tbody>
</table>

Table 12: Coronary intra-observer agreements for observer 2.

<table>
<thead>
<tr>
<th>Observer 1: original / observer 2: MPEG\textsubscript{12}</th>
<th>Observer 1: original / observer 2: MLOT\textsubscript{12}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coro. branch</td>
<td>N\textsubscript{patient}</td>
</tr>
<tr>
<td>RCA</td>
<td>18</td>
</tr>
<tr>
<td>LAD</td>
<td>18</td>
</tr>
<tr>
<td>CX</td>
<td>18</td>
</tr>
</tbody>
</table>

Table 11 indicates that the intra-observer agreements of observer 1 are very high when comparing one reading on originals versus one reading on compressed films. The concordance correlation coefficients r\textsubscript{c} for intra-observer agreements are higher than the inter-observer ones and have narrower confidence intervals. The intra-observer agreements of observer 2 vary with the coronary branch and the compression method. The agreement for the LAD is slightly higher than for the RCA and the CX. Furthermore, observer 2 agrees with himself less when one reading is on original and the other on MPEG\textsubscript{12} films. For the three branches, the intra-observer agreement of observer 2 with MPEG\textsubscript{12} is smaller than is agreement with MLOT\textsubscript{12}. This difference is however not significant. Agreements of observer 2 are smaller than agreements of observer 1.

As a conclusion, the intra-observer agreements are very high for observer 1. The intra-observer agreements of observer 2 are globally smaller, especially when one of his reading is made with MPEG\textsubscript{12} films.

### 5. CONCLUSION

This paper examined the diagnostic use of digital ventriculograms and coronary arteriograms after lossy image data compression. We assessed the performance of a standard algorithm, MPEG, and a dedicated algorithm, MLOT. MPEG has been proposed by the International Standardization Organization for multimedia applications and moving pictures. MPEG is widely used in the computer world and many software and hardware implementations are available. It has a major drawback due to its underlying principle: blocking artifacts are introduced by the block decomposition of the image and are made more visible in the angiograms, due to an edge enhancement processing. To overcome this problem, a new algorithm, MLOT, was developed by Philips Research. It enables higher compression ratios and is free of blocking artifacts.
As there is no reliable measure to predict the visual quality of an image judged by an observer, the assessment of the diagnostic quality of compressed angiograms must be conducted. An evaluation study was done at the Lille Cardiology Hospital with the digital angiograms from 18 patients, and with two expert cardiologists. They interpreted the original and the compressed films blindly and separately. Two compression algorithms were applied: MPEG and MLOT, both at a compression ratio of 12:1. The study protocol was original because it was based on the ordinary clinical practice, which implied the use of complete angiograms (hundreds of images per case). A diagnostic questionnaire was used to record the observers’ interpretations on the left ventriculogram and the coronary arteriogram. The statistical analysis was based on measures of observers’ agreement.

It is noteworthy that the observers were not able to detect the presence of compression on any of the reviewed films.

The left ventricle agreements were measured with the kappa statistics. Inter-observer agreements were good or average, and above chance alone. No statistically significant difference were found for the inter-observer and intra-observer agreements with or without compression.

The coronary agreements were measured with a concordance correlation coefficient introduced by Lin [LIN-89]. Inter-observer agreements were good and intra-observer agreements were excellent. No statistically significant difference was found for the intra-observer agreements with or without compression. One of the two observer has slightly lower intra-observer agreements, especially with MPEG.

This study gives promising results about the application of lossy data compression on cardiac angiograms. More studies should confirm these findings with a stronger focus on intra-observer agreement than our protocol allowed. Complementary type of work are:
- apply the same type of protocol (addressing intra-observer agreements) at a different site (other acquisition equipment, other observers),
- assess the influence of viewing parameters such as zoom, edge enhancement, contrast or brightness,
- evaluate the results of quantitative coronary analysis on compressed images.