Comparison and Evaluation of Information-based Measures in Images

Jorge Miguel Silva[1,2] , Diogo Pratas[1,2,3] and Sérgio Matos[1,2]

[1]IEETA, University of Aveiro, Aveiro, Portugal; [2]DETI, University of Aveiro, Aveiro, Portugal; [3] Department of Virology, University of Helsinki, Helsinki, Finland

Purpose of the study

Lossless data compressors and **small Turing machines** can **approximate the quantity of information** present in a digital object.

In this paper, we **describe** and **compare** these approaches of measuring unsupervised probabilistic and algorithmic information on **images (2D)** with different characteristics.

We use the **Normalized Compression (NC)** employing the data compression **PAQ8** and compare it with the **Block Decomposition Method (BDM)** and show some **advantages** and **limitations** of both measures.

Results

To compare **NC** with **BDM**, we performed **three tests that analyzed**:

- Their **robustness** with **increasing rates of random pixel changes** in **paintings**;
- **Their** behavior on **different** types of **images**;
- Their **minimal information bounds**.

Figure 1 shows that when using the same type of normalization, **NC** is **more robust** than **NBDM** (NBDM1) to the **increase in the rate of random pixel edition.**

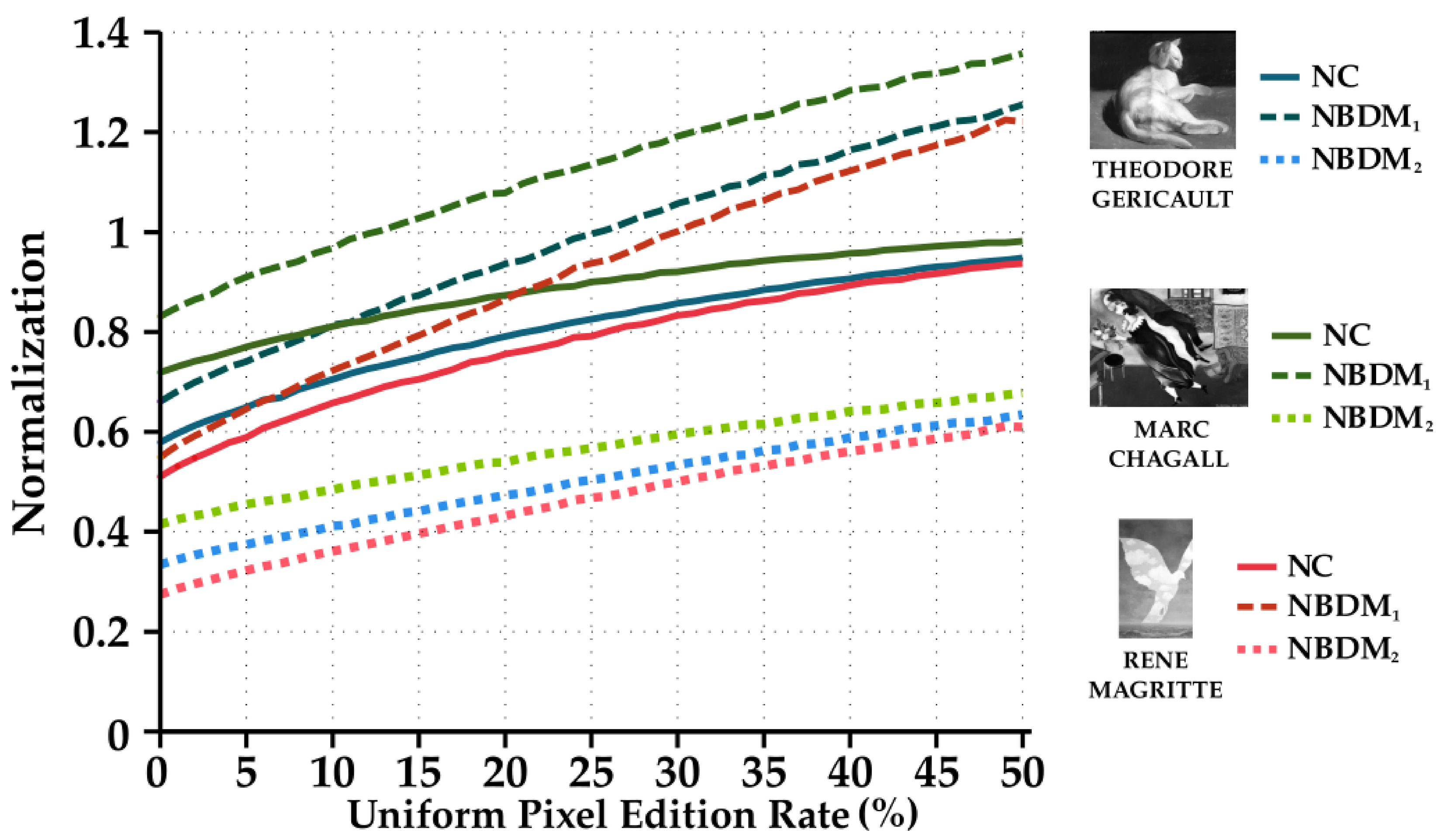

Figure 1. Impact of increasing pseudo-random substitution on information-based measures: NC (approximated using the PAQ8 algorithm) and two BDM normalizations ($NBDM_1$ and $NBDM_2$).

In the **second test**, we applied the **$NBDM_1$** and **NC** to **six** distinct datasets (9 images each) to understand the behavior of these measures for different types of images.

The six datasets were:

- **Artistic images** from (2 datasets);
- **Cellular automata** images;
- **Diabetic retinopathy** images;
- **Chest computed radiography** images;
- **Photographic** images.

Figure 2 shows **different behaviors between NC** and **$NBDM_1$ in images generated by Cellular automata.**
The **BDM** can **ascertain** their **algorithmic nature** and thus attribute them with **minimal value**. This outcome shows the importance of the **BDM** in the **detection of simple output programs embedded into data**.

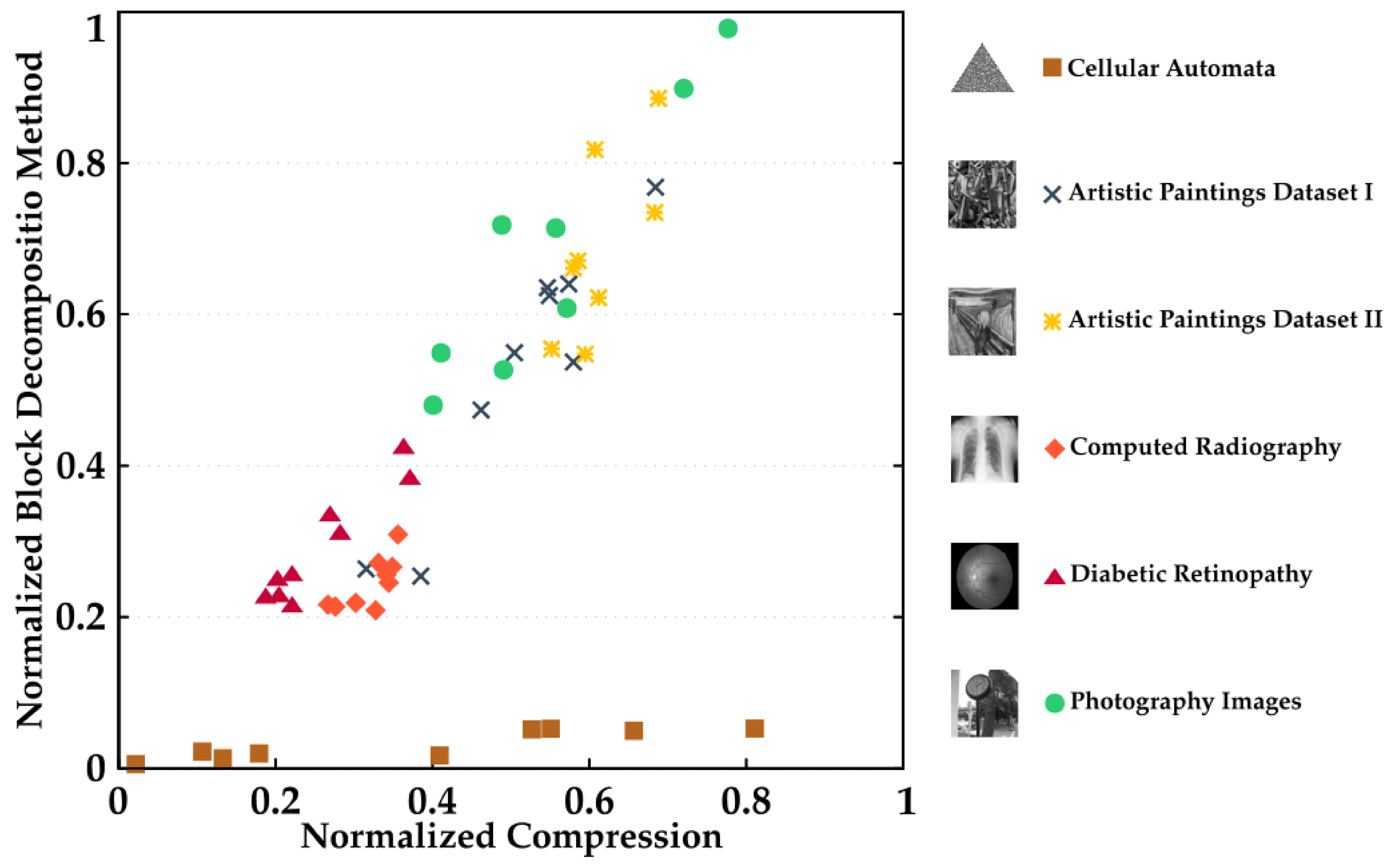

Figure 2. NC and $NBDM_1$ for different types of images.

In the **last test**, we selected one of the **most complex images** identified by the **NBDM** in the last subsection to test if the **BDM** could **accommodate** specific **data alterations**.
Figure 3 shows that after performing a **super-sample image transformation**, the **BDM** was computed for the **original** and the **super-sampled image**.

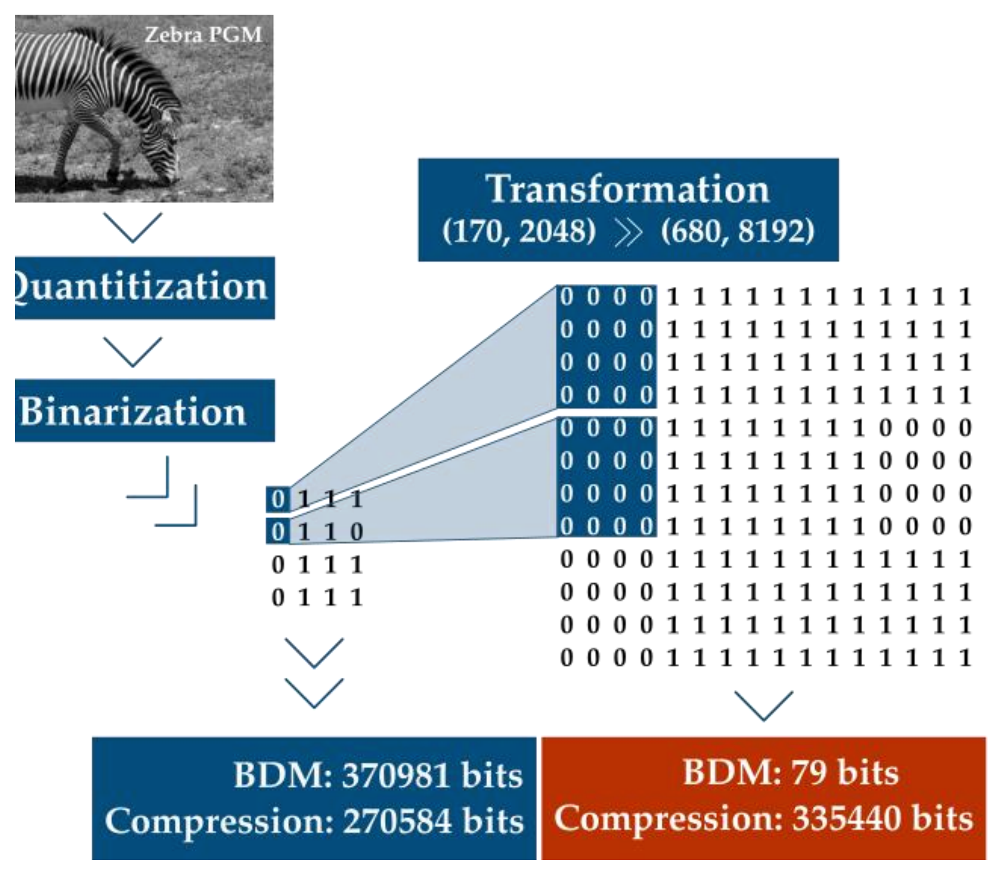

Figure 3. Image transformation pipeline leading to BDM underestimation of the amount of information contained in the transformed object.

The **original image** was measured with **370981 bits**, whereas the **super-sampled** image had only **79 bits**.

This **abrupt decrease** in the complexity value indicates that the **BDM underestimates** the **amount of information** contained **in the object**. This is because **BDM analyses object information in blocks instead of looking at the whole object.**

Conclusions

- **NC** is more **robust to the increment of pixel edition** than **NBDM**.
- **BDM can determine the algorithmic nature of images** created with small programs with simple rules.
- **BDM is not prepared to deal with the information associated with the model's choice, unlike NC**. The NC relies on using a lossless data compressor, bounded by a maximum information channel capacity.
- **There are advantages** and **limitations** of **both measures**. Ranking these measures is not a fair task because they have different characteristics and nature.

REPÚBLICA PORTUGUESA — MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR; FCT Fundação para a Ciência e a Tecnologia; RECPAD'20; universidade de aveiro; ieeta

This work is funded by National Funds through the FCT – Foundation for Science and Technology, by the scholarship SFRH/BD/141851/2018.