

# Accepted Manuscript

## Lossy Compression Approach to Subspace Clustering

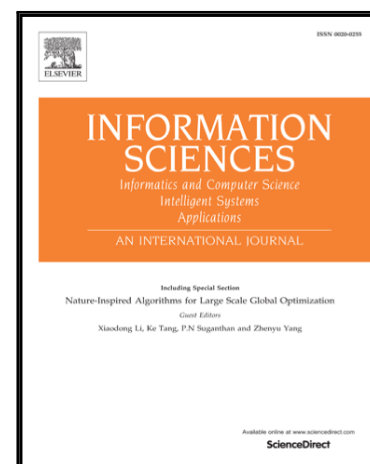
Łukasz Struski, Jacek Tabor, Przemysław Spurek

PII: S0020-0255(16)31162-8  
DOI: [10.1016/j.ins.2017.12.056](https://doi.org/10.1016/j.ins.2017.12.056)  
Reference: INS 13350

To appear in: *Information Sciences*

Received date: 3 October 2016  
Revised date: 27 December 2017  
Accepted date: 29 December 2017

Please cite this article as: Łukasz Struski, Jacek Tabor, Przemysław Spurek, Lossy Compression Approach to Subspace Clustering, *Information Sciences* (2017), doi: [10.1016/j.ins.2017.12.056](https://doi.org/10.1016/j.ins.2017.12.056)



This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

# Lossy Compression Approach to Subspace Clustering

Lukasz Struski\*, Jacek Tabor, Przemysław Spurek

*The Faculty of Mathematics and Computer Science,  
Jagiellonian University, Łojasiewicza 6, 30-348 Kraków, Poland.*

---

## Abstract

We present a novel subspace clustering algorithm **SuMC** (Subspace Memory Clustering) based on information theory, MDLP (Minimal Description Length Principle) and lossy compression. **SuMC** simultaneously solves two fundamental problems of subspace clustering: determination of the number of clusters and their optimal dimensions.

Although **SuMC** requires only two parameters: data compression ratio  $r$  and a number of bits that are used to code a single scalar, the optimal value of compression ratio can be estimated by the Bayesian information criterion (BIC).

We verified that in typical tasks of clustering, image segmentation and data compression, we obtain either better or comparable results to the leading methods of subspace clustering.

*Keywords:* subspace clustering, projected clustering, minimum description length principle, entropy.

---

## 1. Introduction

Clustering techniques have been extensively studied for years in areas such as statistics [6], pattern recognition [10], big data [36, 37] and machine learning. However, most clustering algorithms do not work efficiently in higher dimen-

---

\*Corresponding author

*Email addresses:* lukasz.struski@uj.edu.pl (Łukasz Struski), jacek.tabor@uj.edu.pl (Jacek Tabor), przemyslaw.spurek@uj.edu.pl (Przemysław Spurek)

Download English Version:

<https://daneshyari.com/en/article/6856681>

Download Persian Version:

<https://daneshyari.com/article/6856681>

[Daneshyari.com](https://daneshyari.com)