

# IMPORTANT FEATURE SELECTION BY SPARSE PCA AND APPLICATION IN AUTOMATED CATEGORIZATION

S. Damla Ahipaşaoğlu, Ngai-Man Cheung, Peter Richtárik and Martin Takáč  
 Ahipasaoglu@sutd.edu.sg, N.Man.Cheung@gmail.com, Peter.Richtarik@ed.ac.uk, Takac.MT@gmail.com



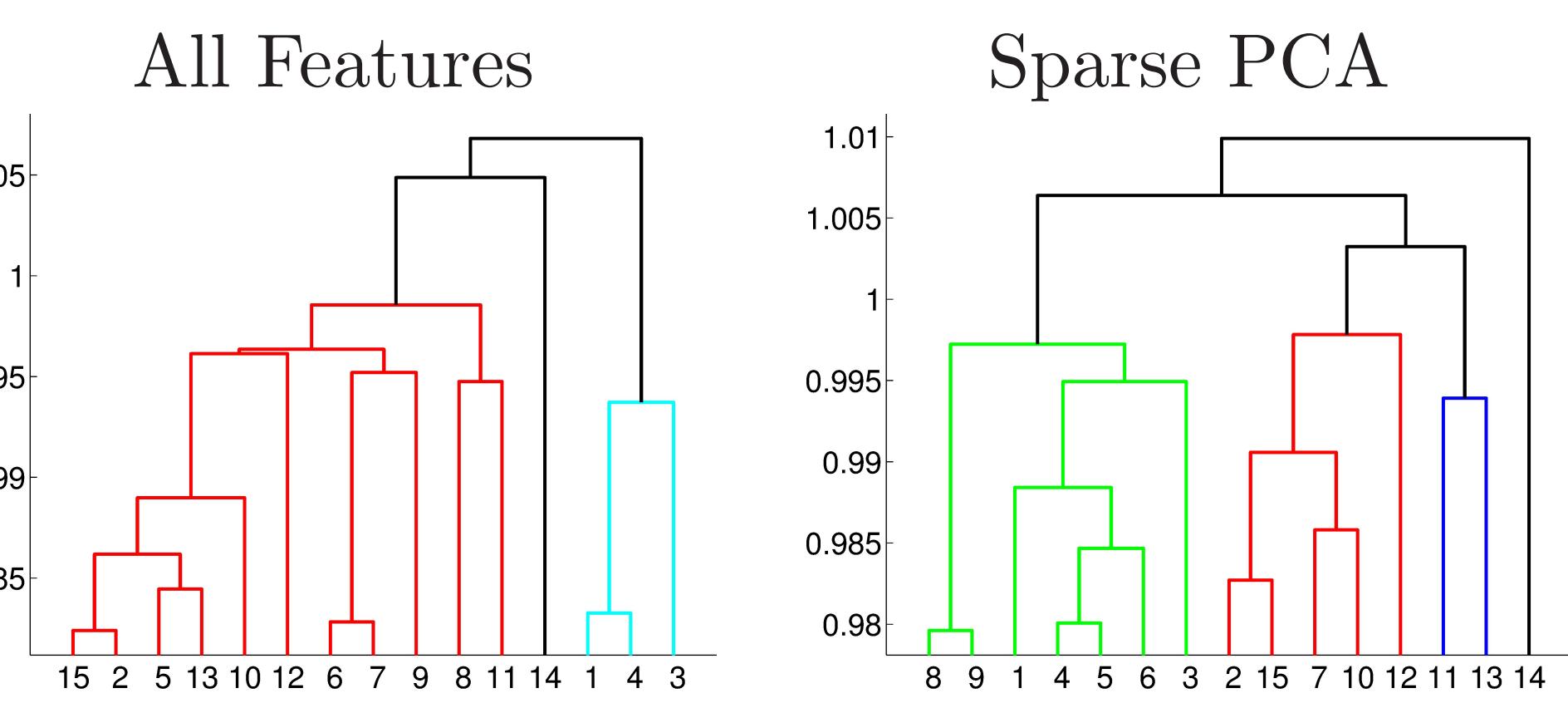
## 1. INTRODUCTION

**Feature** - Interesting points (corners, edges, ...) on the picture is equipped with 64 numbers (descriptor)

**Drawback** - Often unrelated features are extracted  
 - belonging to noise (people staying in front of building, trees, ...)

**Goal of Good Feature Selection:** Identify features, which helps to distinguish objects of different categories

## 2. HIERARCHICAL CATEGORIZAT.



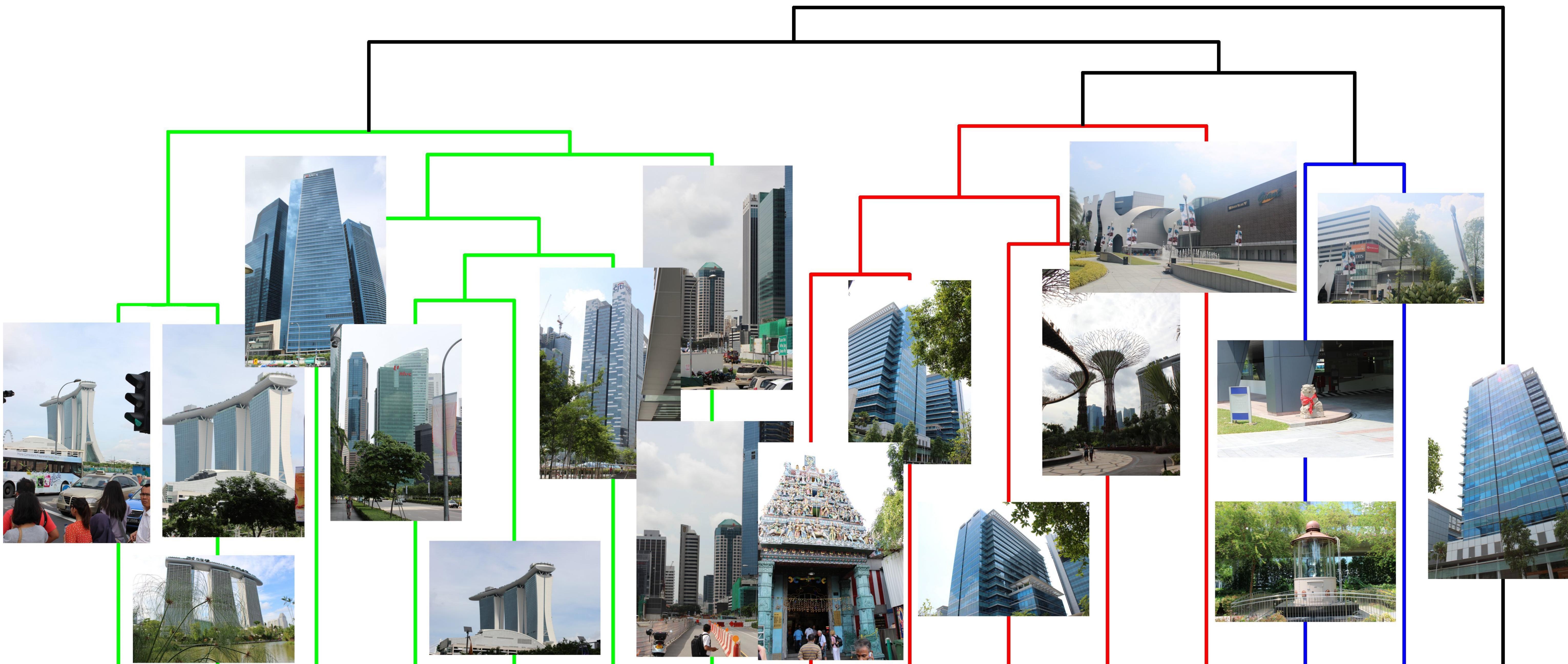
Features selected by Sparse PCA make it easier and more robust in clustering

## 3. SPARSE PCA IS ABLE TO SELECT IMPORTANT FEATURES



Top - original image, Middle - image with all features (MSRE), Bottom - features selected by sparse PCA

## 4. CATEGORIES IDENTIFIED BY CLUSTERING USING A SUBSET OF FEATURES (SELECTED BY SPARSE PCA)



## 5. FINAL CATEGORIES VS. GROUND TRUTH

Identified Categories	Ground Truth Categories														
	0	0	<b>20</b>	0	0	0	0	0	0	0	0	0	0	0	0
<b>15</b>	0	0	<b>19</b>	0	<b>2</b>	0	0	0	0	0	0	0	0	0	0
0	0	0	0	<b>3</b>	0	0	0	0	0	0	<b>3</b>	0	0	0	0
0	0	0	0	0	<b>19</b>	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	<b>5</b>	0	0	0	0	0	0	0	0	0
0	<b>8</b>	0	0	0	0	0	<b>3</b>	<b>8</b>	<b>7</b>	<b>18</b>	<b>6</b>	<b>46</b>	<b>14</b>	0	0
0	0	0	0	0	<b>2</b>	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	<b>3</b>	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	<b>8</b>	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	<b>2</b>
0	0	0	0	0	0	0	0	0	0	0	0	<b>7</b>	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	<b>17</b>	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	<b>2</b>	0	0

Total 265 pictures in 15 categories.