IMAGE FORGERY DETECTION

Context:
Recently, there have been many cases of fake footage of criminal evidences.Today's technology allows digital media to be altered and manipulated in ways that were
simply impossible 20 years ago.

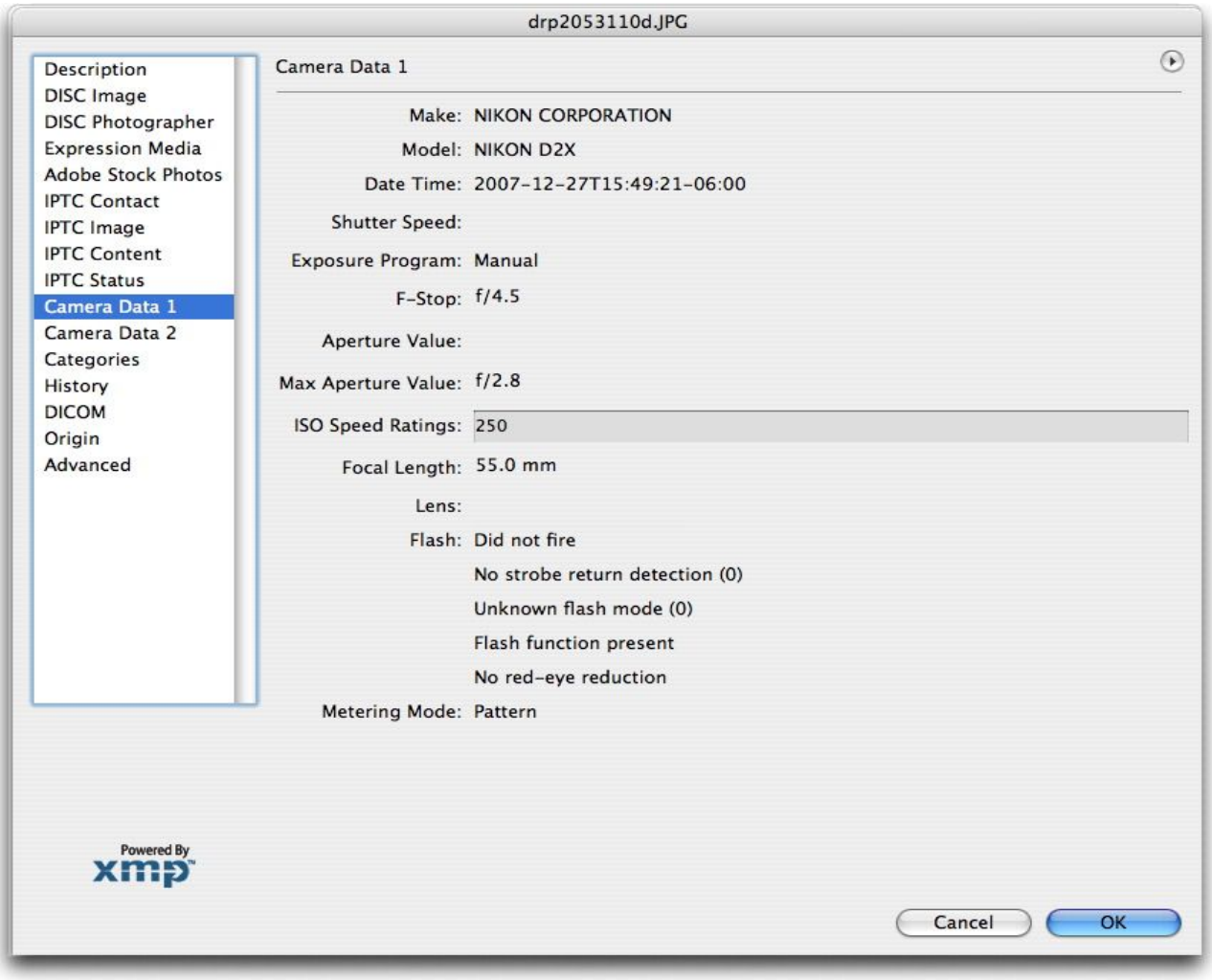
Motivation:
Although, metadata is a way for finding its validity,but it can be easily falsified. So,my motive would be to build an algorithm that identifies which camera model captured an image by using traces intrinsically left in the image. Helping to solve this problem would have a big impact on the verification of evidence used in criminal and civil trials and even news reporting.

Prior Art:
Digital watermarking has been proposed but this has an disadvantage that a watermark must be inserted at the time of recording, which would limit this approach to specially equipped digital cameras.
Also, many machine learning solutions have been proposed in the past: least-squares estimates of a camera's color demosaicing filters as classification features, co-occurrences of pixel value prediction errors as features that are passed to sophisticated ensemble classifiers, and using CNNs to learn camera model identification features. However, this is a problem yet to be sufficiently solved

Datasets:
The dataset is already available at Kaggle but the file is too large (~9GB) so I cannot include it here.

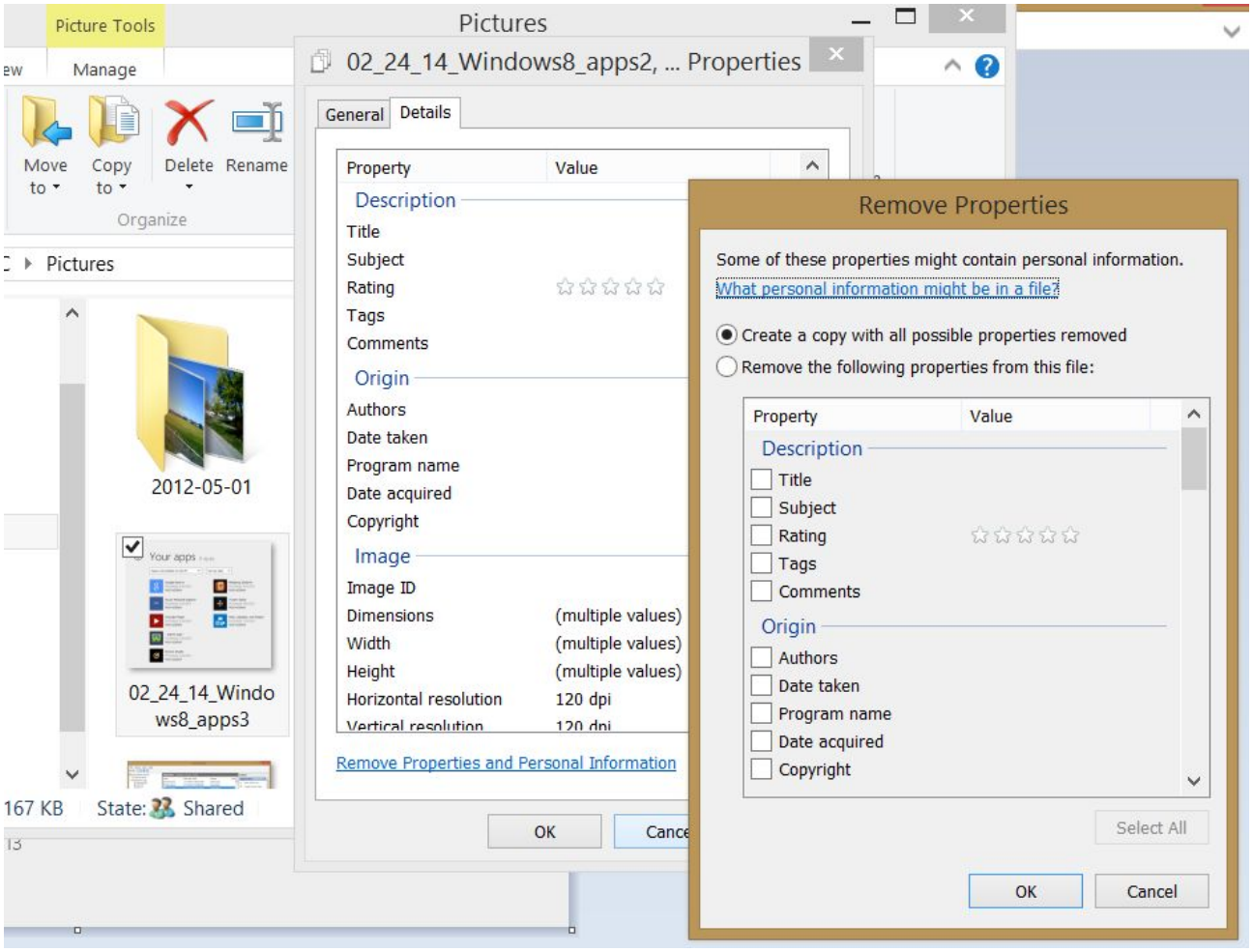
Implementation:
We will find correlation between different features in the metadata but since the metadata can easily be spoofed, we will also try to classify the orientation of pixels, relation between them, the change in color values, color ranges and gamut information.
Different cameras have different image sensors and every sensor is unique in the way it captures light and translates it into pixels. The type of lens used also affects the pixel values.
In a single frame, we can try to identify if the color ranges of all the pixels fall under an individual sensors or multiple sensors are being used and hence detect forgery. Other parameters can also be used in this.
Additionally, we know that photoshopped images tend to have different pixel crispness and resolution is then calculated overall using aggregate functions to normalise. We can detect the correlation between different pixels and determine if they were originally captured or calculated during normalisation

process. After determining the possible values of all the pixels' originality, we can see the image as a whole and check if the patterns match and the image is a result of resolution enhancement or is actually forged (if there are more aggregated pixels in one part of the image as compared to other parts).



Original metadata of an image

IMAGE FORGERY DETECTION



Spoofing metadata of an images