Source Camera Identification Forensics Based on Wavelet Features

Bo Wang, Yiping Guo,
Xiangwei Kong, Fanjie Meng

Dalian University of Technology, China

Outline

- Introduction
- Image features based identification
- Kharrazi's method
- Our method
- Experimental results and conclusions

Introduction

Source Camera Identification:

Identifying the source camera of a digital photograph

Used for:

Establishing the origin of legal photographic evidence

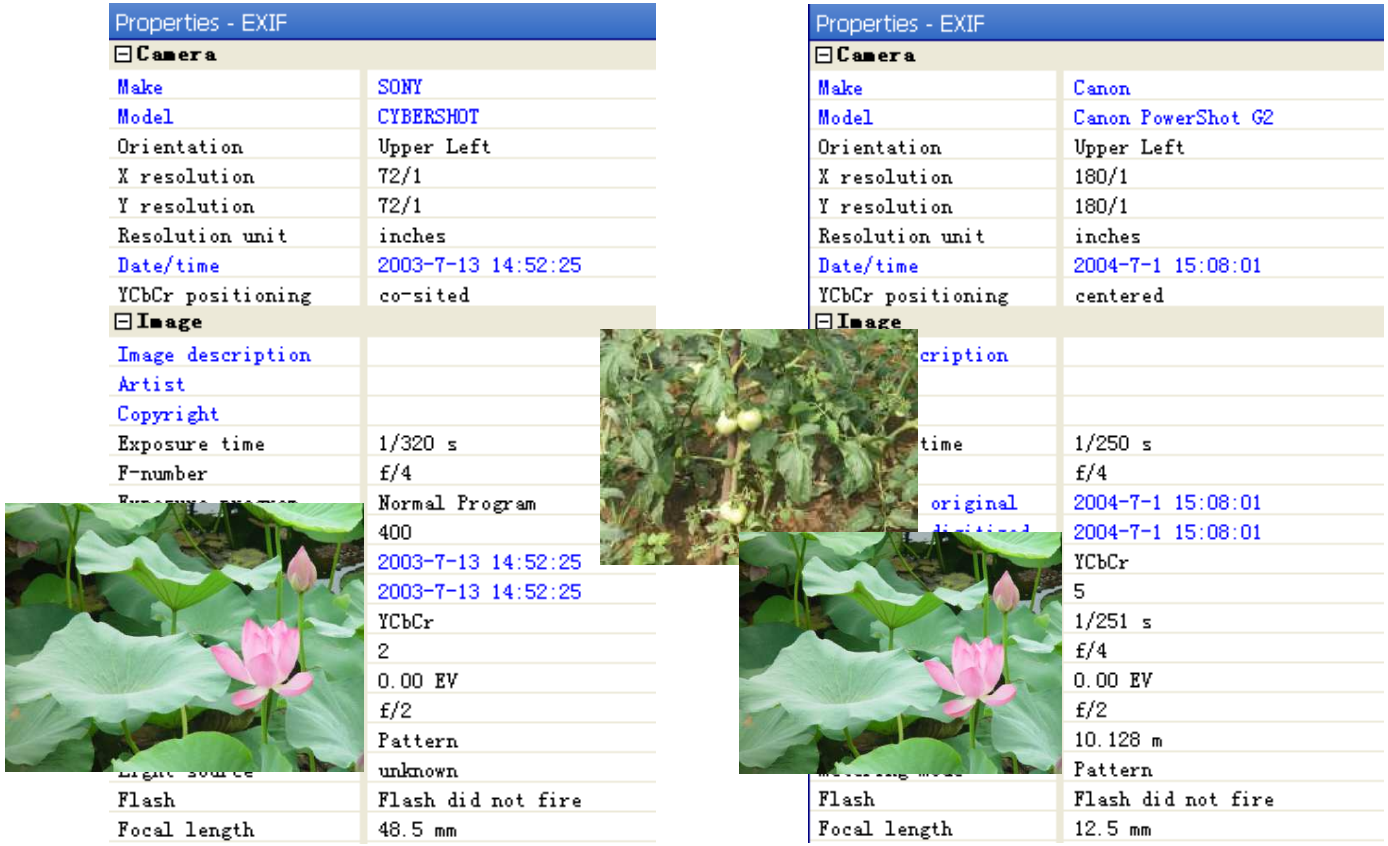
Active and Passive Identification

- **Active Identification**
  - Embed watermarks

  No watermarks in most of digital photographics
- **Passive Identification**
  - Do not need embed any information
  - Only using image data

  Our method is a passive identification

Using EXIF for Identification



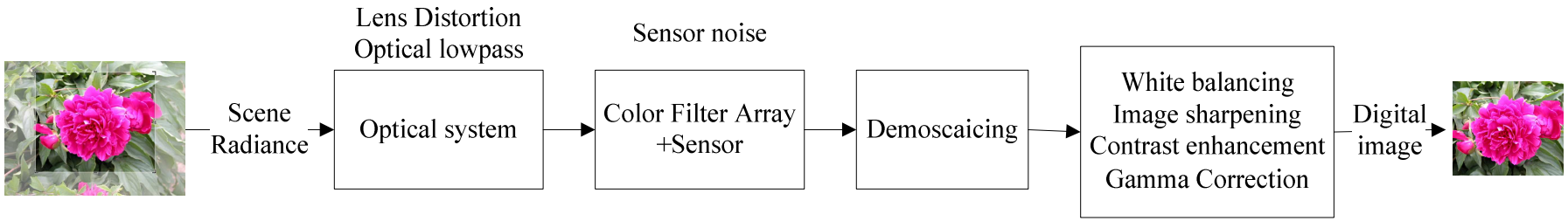
Properties - EXIF	
Camera	
Make	SONY
Model	CYBERSHOT
Orientation	Upper Left
X resolution	72/1
Y resolution	72/1
Resolution unit	inches
Date/time	2003-7-13 14:52:25
YCbCr positioning	co-sited
Image	
Image description	
Artist	
Copyright	
Exposure time	1/320 s
F-number	f/4
	Normal Program
	400
	2003-7-13 14:52:25
	2003-7-13 14:52:25
	YCbCr
	2
	0.00 EV
	f/2
	Pattern
	unknown
Flash	Flash did not fire
Focal length	48.5 mm

Properties - EXIF	
Camera	
Make	Canon
Model	Canon PowerShot G2
Orientation	Upper Left
X resolution	180/1
Y resolution	180/1
Resolution unit	inches
Date/time	2004-7-1 15:08:01
YCbCr positioning	centered
Image	
cription	
time	1/250 s
	f/4
original	2004-7-1 15:08:01
	2004-7-1 15:08:01
	YCbCr
	5
	1/251 s
	f/4
	0.00 EV
	f/2
	10.128 m
	Pattern
Flash	Flash did not fire
Focal length	12.5 mm

- **which one is its original EXIF? The left one.**
- **The EXIF of the right one is replaced by another image.**

Image Features Based Identification

- **Imaging pipeline in digital cameras**

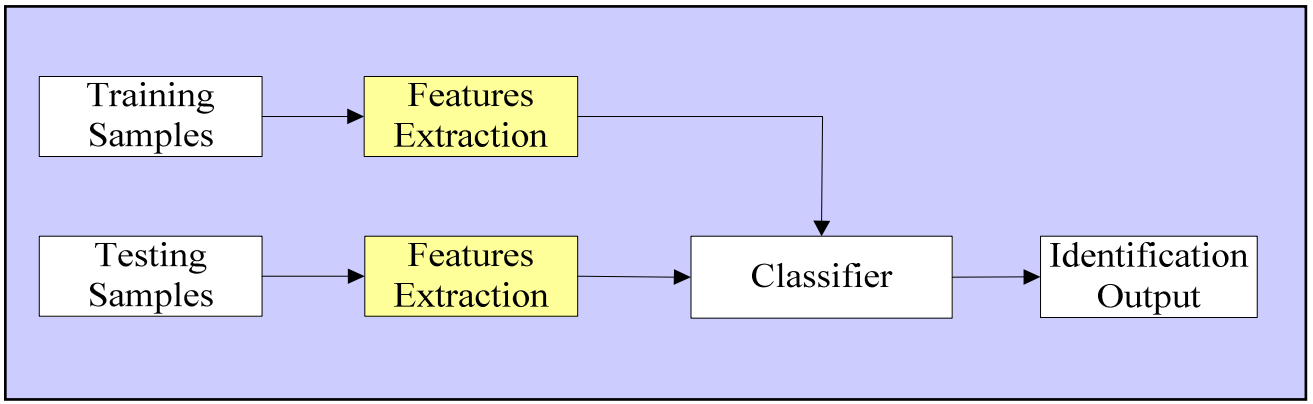


Differences in the processing details of each stage among various models of digital cameras

Differences of image features in the output images from cameras of different models

Kharrazi's Method

- Polytechnic University, Brooklyn, NY,USA: Mehdi Kharrazi, Husrev T. Sencar, Nasir Memon
- Using Pattern Recognition



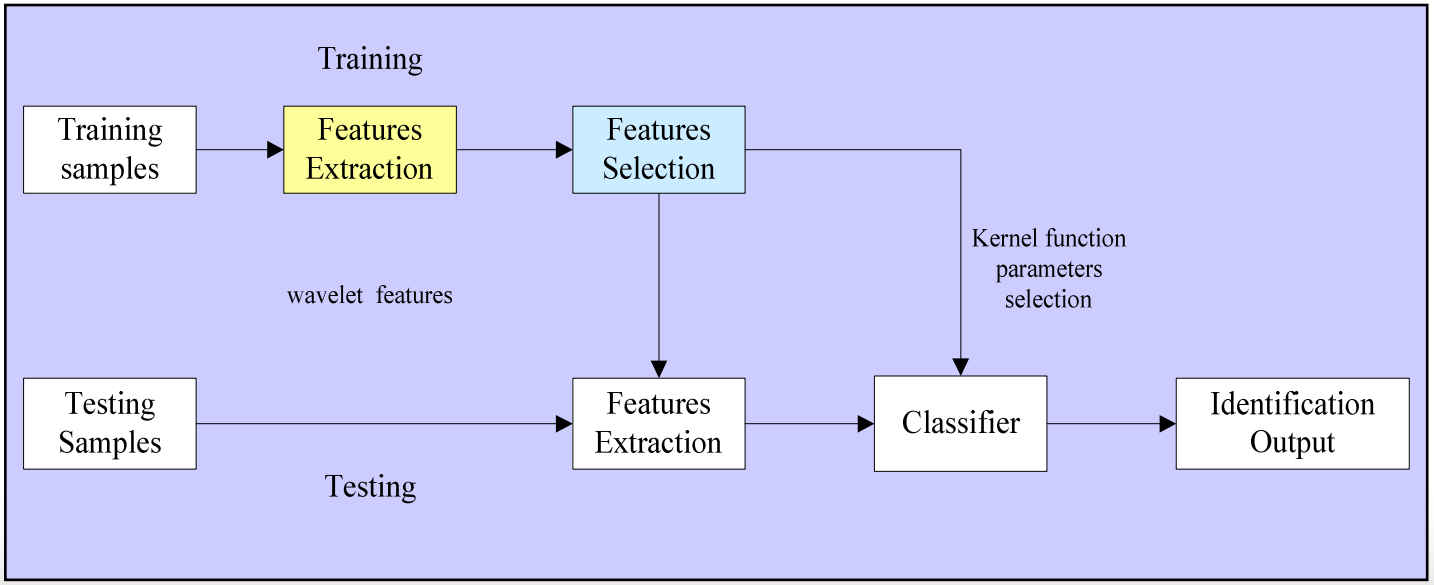
- Image Features: color features,IQM features, mean of wavelet coefficients

Can we do better?

- Shortage of Kharrazi's method
- **Identification accuracy is not reliable**
- Why?
- **Image Features used are not effective**
- What we do?
- **Extract more effective features**

Our method

- Features Extraction
- Features Selection
- Classification



Wavelet Features

- Higher-order wavelet statistics

➢ Statistics of linear prediction of wavelet coefficients

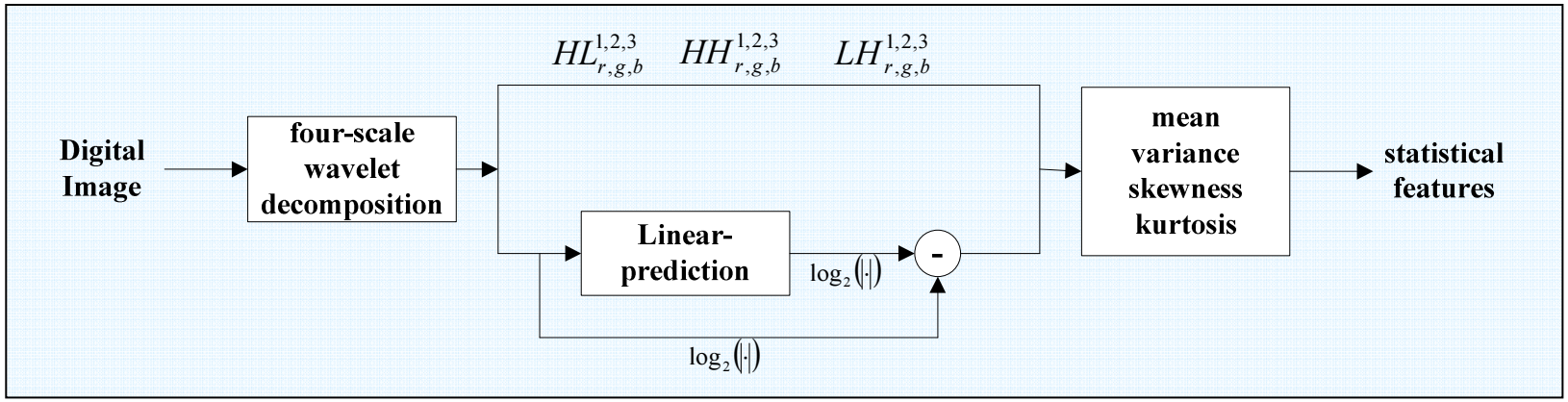
  – **A kind of filter operation in wavelet domain**
  – **Less dependence on image content**

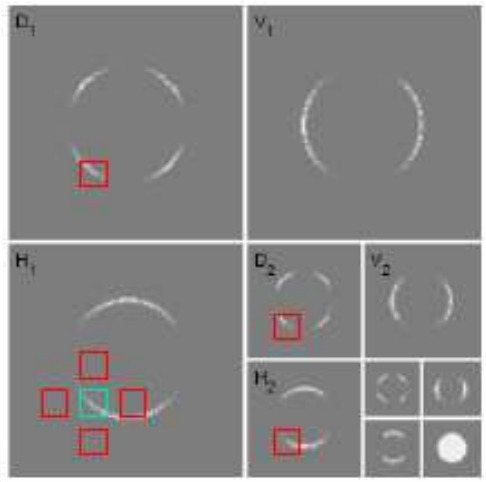
- Wavelet Coefficient Co-occurrence statistics

➢ Distances of co-occurrence matrices in the same orientation between different scales

Wavelet features → Differences in impact of imaging pipelines on wavelet domain

Higher-order Wavelet Features



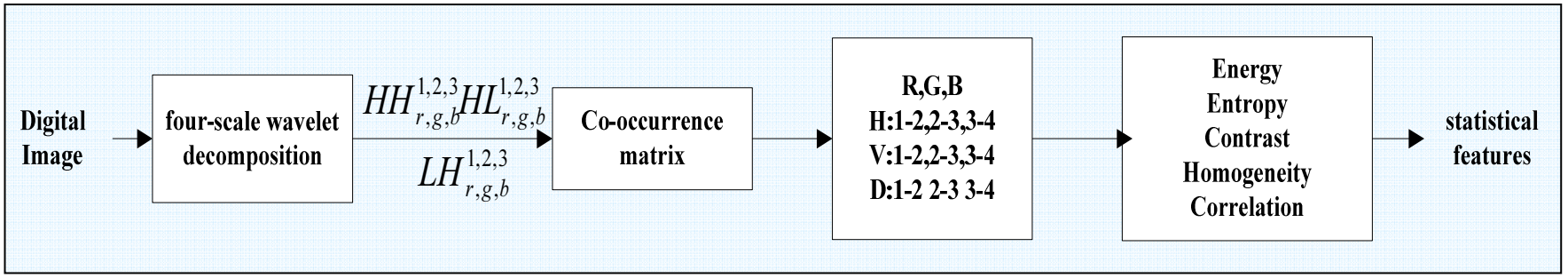


$$\begin{aligned}\left|V_i^g(x,y)\right| &= \omega_1\left|V_i^g(x-1,y)\right| + \omega_2\left|V_i^g(x+1,y)\right| + \omega_3\left|V_i^g(x,y-1)\right| \\ &+ \omega_4\left|V_i^g(x,y+1)\right| + \omega_5\left|V_{i+1}^g(x/2,y/2)\right| + \omega_6\left|V_i^g(x,y)\right| \\ &+ \omega_7\left|D_{i+1}^g(x/2,y/2)\right| + \omega_8\left|V_i^r(x,y)\right| + \omega_9\left|V_i^b(x,y)\right|\end{aligned}$$

$$\vec{\upsilon} = Q\vec{\omega} \qquad E(\vec{\omega}) = \left[\vec{\upsilon} - Q\vec{\omega}\right]^2 \qquad \frac{dE(\vec{\omega})}{d\vec{\omega}} = 2Q^T(\vec{\upsilon} - Q\vec{\omega})$$

$$\vec{\omega} = (Q^TQ)^{-1}Q^T\vec{\upsilon} \qquad \vec{p} = \log(\vec{\upsilon}) - \log\left(\left|Q\vec{\omega}\right|\right)$$

Wavelet Coefficient Co-occurrence Statistics



$$DC(V_i^c) = CV_i^c - CV_{i+1}^c$$

$$DC(H_i^c) = CH_i^c - CH_{i+1}^c$$

$$DC(D_i^c) = CD_i^c - CD_{i+1}^c$$

CV_i^c CH_i^c CD_i^c : vertical, horizontal, and diagonal subbands' co-occurrence matrices

$$i = 1,2,3,4 \quad c = r,g,b$$

$$Energy = \sum_i \sum_j DC^2[i,j]$$

$$Entropy = -\sum_i \sum_j DC[i,j] \log_2 DC[i,j]$$

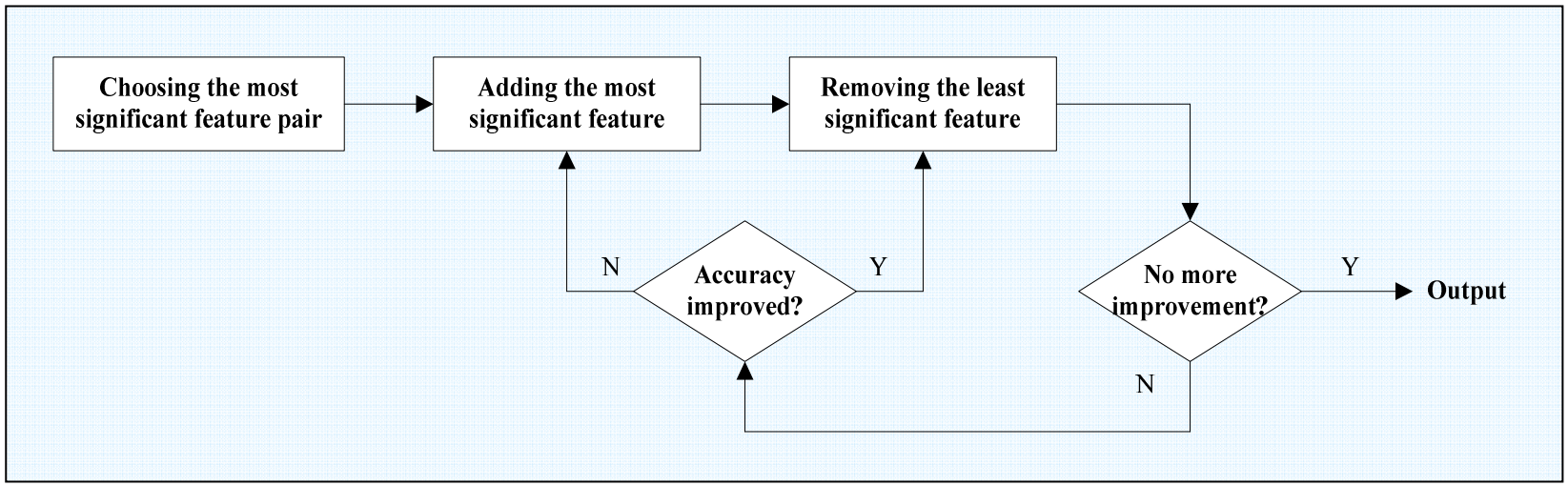
$$Contrast = \sum_i \sum_j (i-j)^2 DC[i,j]$$

$$Homogeneity = \sum_i \sum_j \frac{DC(i,j)}{1+|i-j|}$$

$$Correlation = \frac{\sum_i \sum_j (i-\mu_i)(j-\mu_j) DC[i,j]}{\sigma_i \sigma_j}$$

Feature Selection and Classification

- Sequential Forward Feature Selection (SFFS)



- Support Vector Machine (SVM)
  - C-support vector classification with non-linear RBF kernel

Experiment

- Experiment samples and parameters

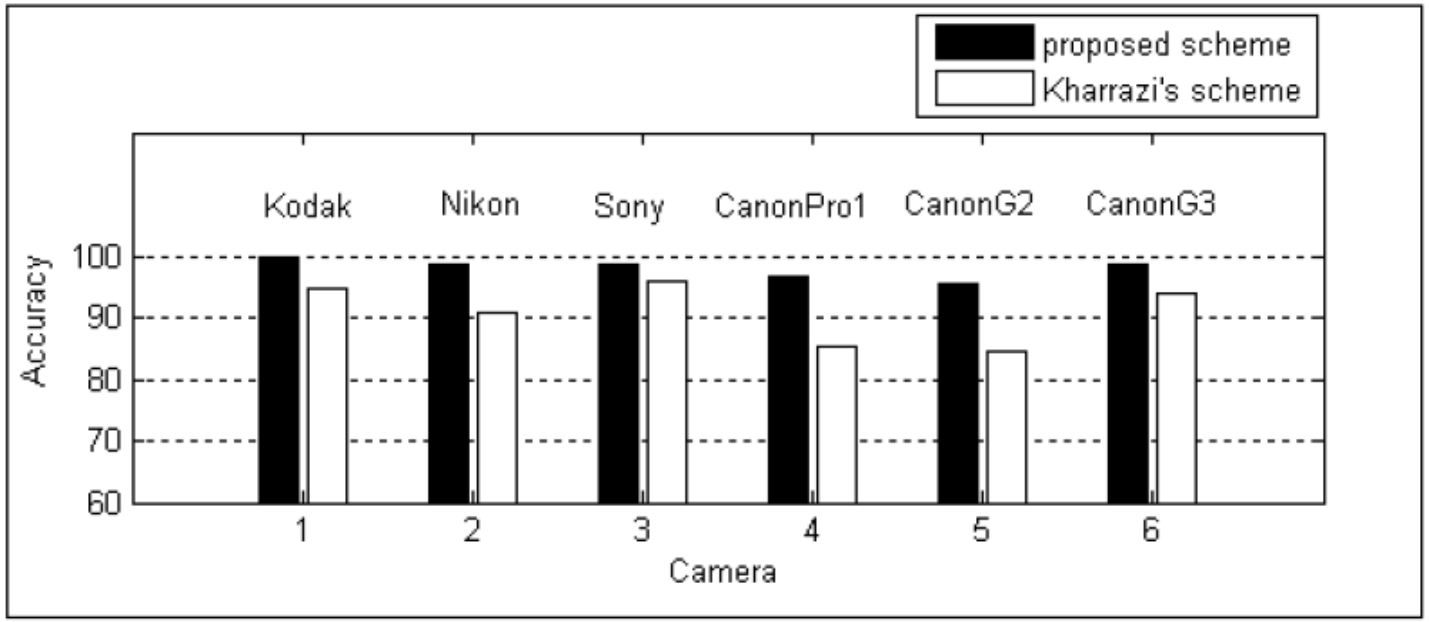
Cameras	Camera Parameters		Sample image parameters	
	Sensor	Max resolution	Image resolution	Image format
Kodak DC290	Unspecified CCD	2240*1500	2240*1500	JPEG
Nikon E5700	2/3 inch CCD	2560*1920	1600*1200	JPEG
Sony DSC-F828	2/3 inch CCD	3264*2448	1280*960	JPEG
Canon PowerShot Pro1	2/3 inch CCD	3264*2448	1024*768	JPEG
Canon PowerShot G2	1/1.8 inch CCD	2272*1704	1024*768 1600*1200 2272*1704	JPEG
Canon PowerShot G3	1/1.8 inch CCD	2272*1704	2272*1704	JPEG

Experiment result of our method

- Confusion matrix

Camera	Kodak	Nikon	Sony	CanonPro1	CanonG2	CanonG3	Accuracy
Kodak DC290	**150**	0	0	0	0	0	100%
Nikon 5700	0	**148**	0	2	0	0	98.7%
Sony DSC-F828	0	2	**148**	0	0	0	98.7%
Canon PowerShot Pro1	0	0	1	**145**	4	0	96.7%
Canon PowerShot G2	0	0	0	3	**143**	4	95.3%
Canon PowerShot G3	0	0	0	0	2	**148**	98.7%

Comparison with Kharrazi's method



Camera	Kodak	Nikon	Sony	CanonPro1	CanonG2	CanonG3	Accuracy
Kharrazi's method	94.7%	91.3%	96.3%	85.3%	84.7%	93.3%	90.9%
Our method	100%	98.7%	98.7%	96.7% ↑11.4%	95.3% ↑10.6%	98.7% ↑ 5.4%	98.2%

Conclusions

1. Introduce feature based source camera identification
2. Discuss a classic feature based identification method
3. Give a new source camera identification method based on wavelet features

Thank you!