Improving the Visible with the Invisible
Incorporating Near-Infrared Cues in Computational Photography and Computer Vision

Sabine Süsstrunk
Images and Visual Representation Group (IVRG)
The “Internet of Things”

- Humans generated *more data* in 2009 than in the *previous 5’000* years combined.\(^1\)

---

\(^1\)Dave Evans, Chief Futurist for Cisco Systems, 2010.
The “Internet of Things”

• Humans generated more data in 2009 than in the previous 5’000 years combined.¹

• By 2012, 90% of all data will be video.¹

¹Dave Evans, Chief Futurist for Cisco Systems, 2010.
The “Internet of Things”

- Humans generated *more data* in 2009 than in the *previous 5’000* years combined.¹
- By 2012, *90% of all data* will be video.¹
- *24 hours* of video are uploaded *every minute* on youtube.²

The “Internet of Things”

• Humans generated *more data* in 2009 than in the *previous 5’000* years combined.¹
• By 2012, *90% of all data* will be video.¹
• *24 hours* of video are uploaded *every minute* on youtube.²
• Over *5 billion* photos on flickr.³

¹Dave Evans, Chief Futurist for Cisco Systems, 2010. ²Google, 2010. ³CNN, 2010
The “Internet of Things”

• Humans generated more data in 2009 than in the previous 5’000 years combined.¹
• By 2012, 90% of all data will be video.¹
• 24 hours of video are uploaded every minute on youtube.²
• Over 5 billion photos on flickr.³
• Each month, 2.5 billion photos are uploaded to Facebook.³

¹Dave Evans, Chief Futurist for Cisco Systems, 2010. ²Google, 2010. ³CNN, 2010
The Challenge of Imaging

How do we render, interpret, organize, and communicate all this visual data?
The Challenge of Imaging

How do we render, interpret, organize, and communicate all this visual data?

we = our machines
Fundamental Imaging Problem

• What is the contribution to the pixel value of the object *reflectance* versus the *light source*?
Fundamental Imaging Problem

- What is the contribution to the pixel value of the object *reflectance* versus the *light source*?

A *red* surface under white light?  

or

A white surface under *red* light?
Fundamental Imaging Problem

• What is the contribution to the pixel value of the object *reflectance* versus the *light source*?

  *Two* shades of tile under one light?
  One shade of tile under *two* lights?
Image Formation

\[ p_k : \text{Camera Response} \]

\[ p_k = \int_{\lambda=400nm}^{700nm} E(\lambda) S(\lambda) Q_k(\lambda) d\lambda \]

\[ k = \{B, G, R\} \]

- Light Power \( E(\lambda) \)
- Surface Reflectance \( S(\lambda) \)
- Quantum Efficiency \( Q(\lambda) \)
Scene Understanding

• Separating the effect of illuminant and surface reflectance of the camera response (pixel value) is an *ill-posed* problem.
Scene Understanding

• Separating the effect of illuminant and surface reflectance of the camera response (pixel value) is an *ill-posed* problem.

• To estimate the two, we either make assumptions about the scene...
Scene Understanding

• Separating the effect of illuminant and surface reflectance of the camera response (pixel value) is an *ill-posed* problem.

• To estimate the two, we either make assumptions about the scene...

• ...or we capture more physical information about the scene.
Current Imaging: RGB

Visible Spectrum

400 nm 700 nm

Visible (Red, Green, and Blue: RGB)
Scientific Imaging: Near-Infrared

Near-Infrared

700 nm

1100 nm

Near-infrared (NIR)
Future Imaging: RGB + NIR

Visible Spectrum

Near-Infrared

Visible (Red, Green, and Blue: RGB)

Near-infrared (NIR)
Benefits of RGB + NIR

- We can **exploit** the correlation and de-correlation of visible and NIR image frequencies and intensities to:
  - **enhance** the visual quality of images/videos.
  - **extract** more accurate information about the scene.
Outline

• Acquisition of RGB + NIR
• Enhance imaging:
  - Haze removal
  - Skin smoothing
• Extract more information about the scene:
  - Illuminant and shadow detection
  - Material-based object segmentation
  - Scene recognition
• Conclusion
Acknowledgment

- Joint work with
  - Nathalie Barbuscia
  - Matthew Brown
  - Damien Firmenich
  - Clément Fredembach
  - Arthur Germain
  - Yue M. Lu
  - Neda Salamati
  - Zahra Sadeghipoor
  - Lex Schaul
  - Daniel Tamburrino

- Funding
  - Swiss National Science Foundation
  - Xerox Foundation
Outline

• **Acquisition of RGB + NIR**
• Enhance imaging:
  - Haze removal
  - Skin smoothing
• Extract more information about the scene:
  - Illuminant and shadow detection
  - Material-based object segmentation
  - Scene recognition
• Conclusion
Human vision and the spectrum

Gamma rays  X-rays  UV  Infrared  Radio waves

Visible Light

0.0001 nm  0.01 nm  10 nm  1000 nm  0.1 cm  1 cm  1 m  100 m
Human vision and the spectrum

- Gamma rays
- X-rays
- UV
- Infrared
- Radio waves

Visible Light

Sensitivity vs. Wavelength
- S-cones
- M-cones
- L-cones

Wavelength: 400 nm to 700 nm
The sensitivity of silicon
Camera Design

Bayer Color Filter Array (CFA)
Camera Sensitivity

Bayer Color Filter Array (CFA)
Camera Sensitivity

Near Infrared Blocking Filter (Hot Mirror)

Quantum Efficiency

400 nm  700 nm  1100 nm  wavelength
All digital cameras are inherently sensitive to NIR...
All digital cameras are inherently sensitive to NIR...

...if we remove the hot mirror!
Camera construction

C. Fredembach and S. Süsstrunk, Colouring the near infrared, IS&T/SID 16th Color Imaging Conference, 2008. MERL Best Student Paper Award
Camera construction

...or you buy one on-line!

C. Fredembach and S. Süsstrunk, Colouring the near infrared, *IS&T/SID 16th Color Imaging Conference*, 2008. *MERL Best Student Paper Award*
Capturing visible images

Visible Image
Capturing NIR images

NIR Image = NIR Image
Capturing full spectrum images
Full Spectrum Image
Current Work

- **Simultaneous** capture of RGB and NIR on a single sensor.
“Multispectral” Color Filter Array

- Exploit spatial and spectral correlation to acquire 4-channel RGB+NIR:
“Multispectral” Color Filter Array

- Exploit spatial and spectral correlation to acquire 4-channel RGB+NIR:

50 RGB/NIR image pairs
“Multispectral” Color Filter Array

- The new color filter array (CFA) for 4-channel RGB +NIR with no hot mirror:

Original RGB
Reconstructed RGB

Av. CPSNR = 34.30 (38.47 for RGB only)
Original NIR
Reconstructed NIR

Av. PSNR = 34.30
Outline

• Acquisition of RGB + NIR

• *Enhance imaging:*
  - *Haze removal*
  - Skin smoothing

• Extract more information about the scene:
  - Illuminant and shadow detection
  - Material-based object segmentation
  - Scene recognition

• Conclusion
EPFL

= Ecole Polytechnique Fédéral de Lausanne
= Swiss Federal Institute of Technology, Lausanne
Normal view...
Enhanced view
Haze Removal

*Light scattering* in the atmosphere produces haze (Raleigh’s law):

\[ s_{\text{particle}} < \frac{\lambda}{10} : \quad E_s \propto \frac{E_0}{\lambda^4} \]
Haze Removal

**Light scattering** in the atmosphere produces haze (Raleigh’s law):

\[ s_{\text{particle}} < \frac{\lambda}{10} : \quad E_s \propto \frac{E_0}{\lambda^4} \]
Intuition

• The *high frequencies* of the visible image can be *exchanged* by the high frequencies of the NIR image.
Algorithm

- Multi-resolution analysis and fusion of both visible luminance $V$ and NIR intensities $N$ (criterion: contrast):

Algorithm

- Multi-resolution analysis and fusion of both visible $V$ and NIR $N$ images.

- Analysis Criterion: **Weber Contrast**

\[ C = \frac{Y - Y_s}{Y_s} \]

Algorithm

- Multi-resolution analysis and fusion of both visible $V$ and NIR $N$ images.

Analysis Criterion:

$L.\ Schaul,\ C.\ Fredembach,\ and\ S.\ Süsstrunk,\ Color\ image\ de-hazing\ using\ the\ near-infrared,\ IEEE\ International\ Conference\ on\ Image\ Processing\ (ICIP),\ 2009.$
Synthesis

- Keep the visible low frequency information, and then retain the highest contrast from either the visible or the NIR image.

\[ F_0 = V_n^a \prod_{k=1}^{n} (\max(V_k^d, N_k^d) + 1) \]

Results

Visible  USM  Gaussian  Bilateral
Results

Results

Original

De-hazing using NIR

Results


High Dynamic Range

High Dynamic Range

Outline

- Acquisition of RGB + NIR
- Enhance imaging:
  - Haze removal
  - *Skin smoothing*
- Extract more information about the scene:
  - Illuminant and shadow detection
  - Material-based object segmentation
  - Scene recognition
- Conclusion
Penetration of radiation in skin

- **Absorption** and **scattering** are inversely proportional to wavelength.

Intuition

- If NIR penetrates deeper into skin, then *surface* imperfections are *less* visible.
Intuition

- Unwanted skin imperfections (freckles, pores, warts, wrinkles) are usually \textit{small}.
Algorithm

\[ T_{\text{RGB-YCC}} \rightarrow \text{BF} \]

\text{BF} = \text{Bilateral Filter}
Algorithm

RGB $\xrightarrow{T_{\text{RGB-YCC}}} Y$ $\xrightarrow{\text{BF}} Y_{\text{detail}}$

NIR $\xrightarrow{\text{BF}}$ NIR$_{\text{detail}}$ $\xrightarrow{\text{BF}}$ NIR$_{\text{base}}$

BF = Bilateral Filter
Algorithm

RGB

T_{RGB-YCC}

BF

Y

Y_{detail}

Y_{base}

NIR

BF

NIR_{detail}

NIR_{base}

Y_{smooth}

RGB_{smooth}

BF=Bilateral Filter
Skin Smoothing

Visible

NIR

Visible + NIR

Skin Smoothing

Camera Museum in Vevey

Camera Museum in Vevey

Visible

Visible + NIR

Preference: 25.4% (1,235)  Preference: 74.6% (3,623)

S. Süsstrunk, C. Fredembach, and D. Tamburrino, Automatic skin enhancement with visible and near-infrared image fusion, ACM Multimedia, 2010. Best Demo Paper Award
Dark Flash Photography

Video Re-Lighting

Outline

• Acquisition of RGB + NIR
• Enhance imaging:
  - Haze removal
  - Skin smoothing
• **Extract more information about the scene:**
  - *Illuminant and shadow detection*
  - Material-based object segmentation
  - Scene recognition
• Conclusion
Recall

- Separating the effect of illuminant and surface reflectance of the camera response (pixel value) is an *ill-posed* problem.

\[
p_k = \int_{\lambda=400nm}^{700nm} E(\lambda) S(\lambda) Q_k(\lambda) d\lambda
\]

![Light Power](image1)
![Surface Reflectance](image2)
![Quantum Efficiency](image3)
Intuition

- **Doubling** the spectral bandwidth considered (400-1100nm) increases estimation accuracy for color constancy algorithms.
Light Sources in the visible spectrum

Normalized power

- Daylight
- Skylight
- Flashlight
- Incandescent
- Fluorescent

wavelength
Light Sources (visible + NIR)

- **Incandescent**
- **Flash**
- **Daylight**
- **Skylight**
- **Fluorescent**

Normalized power across different wavelengths for various light sources, including visible and near-infrared (NIR) regions.
Light Source Estimation

Overcast Sky
Estimated Color Temperature: 7500K

Sunny Day
Estimated Color Temperature: 5000K

Incandescent Lamp
Estimated Color Temperature: 3000K

Shadow Detection

Outline

• Acquisition of RGB + NIR
• Enhance imaging:
  - Haze removal
  - Skin smoothing
• Extract more information about the scene:
  - Illuminant and shadow detection
  - *Material-based object segmentation*
  - Scene recognition
• Conclusion
Material Reflectances

Yellow Arrows: same color in RGB, different intensity in NIR
Green Arrow: different colors in RGB, same intensity in NIR

Image Segmentation

- **$P_C$: Color segments**
- **$P_H$: Shadow segments**
- **$P_M$: Material segments**
- **$P_H$: Shadow segments**
Intrinsic Image (RGB + NIR)

N. Salamati and S. Süsstrunk, Material-Based Object Segmentation using Near-Infrared Information, IS&T/SID 18th Color Imaging Conference, 2010. *Best Interactive Paper Award*
N. Salamati and S. Süsstrunk, Material-Based Object Segmentation using Near-Infrared Information, IS&T/SID 18th Color Imaging Conference, 2010. *Best Interactive Paper Award*
Comparison with visible only

Visible + NIR

Visible Only

N. Salamati and S. Süsstrunk, Material-Based Object Segmentation using Near-Infrared Information, IS&T/SID 18th Color Imaging Conference, 2010. *Best Interactive Paper Award*
Outline

• Acquisition of RGB + NIR
• Enhance imaging:
  - Haze removal
  - Skin smoothing
• Extract more information about the scene:
  - Illuminant and shadow detection
  - Material-based object segmentation
  - Scene recognition
• Conclusion
Image Database (477 RGB+NIR)
Correlation between RGB and NIR

- 10’000 RGB-NIR pixel pairs, joint entropy:

\[
\begin{align*}
\text{(a) R-G (4.13)} & \quad \text{(b) B-G (4.26)} & \quad \text{(c) R-B (4.60)} \\
\text{(d) R-NIR (5.29)} & \quad \text{(e) G-NIR (5.30)} & \quad \text{(f) B-NIR (5.44)}
\end{align*}
\]
Decorrelation...

- PCA on 477 4-channel images (RGB +NIR):

M. Brown and S. Süsstrunk, Multi-spectral SIFT for Scene Category Recognition, CVPR 2011.
Multi-spectral SIFT

• 477 RGB+NIR images
  - 9 scene categories
• SIFT, GIST, HMAX descriptors.
• 11 random images per class for training, rest for testing.
  - Repeated 10 times
• \textbf{MSIFT} (=pca 4) statistically significantly \textit{outperforms} the others.

MulR-spectral SIFT

- SIFT, RGB+NIR images - 9 scene categories
  - SIFT, GIST, HMAX descriptors.
  - 11 random images per class for training, rest for testing.
  - Repeated 10 times

- **MSIFT** (=pca 4) statistically significantly **outperforms** the others.

M. Brown and S. Süsstrunk, Multi-spectral SIFT for Scene Category Recognition, CVPR 2011.
## Better Features and Classifiers

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Descriptor1</th>
<th>Descriptor2</th>
<th>Fusion type</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>COL$_{r,g,b}$</td>
<td>SIFT$_{l,n}$</td>
<td>EF</td>
<td>84.3 ± 1.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SIFT$_{n}$</td>
<td>LF</td>
<td>87.4 ± 2.7</td>
</tr>
<tr>
<td></td>
<td>COL$_{r,g,b,n}$</td>
<td>SIFT$_{l,n}$</td>
<td>EF</td>
<td>84.1 ± 2.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SIFT$_{n}$</td>
<td>LF</td>
<td>86.2 ± 2.0</td>
</tr>
<tr>
<td></td>
<td>COL$_{p1,p2,p3}$</td>
<td>SIFT$_{l,n}$</td>
<td>EF</td>
<td>84.1 ± 2.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SIFT$_{n}$</td>
<td>LF</td>
<td>86.5 ± 2.4</td>
</tr>
<tr>
<td></td>
<td>COL$_{p1,p2,p3,p4}$</td>
<td>SIFT$_{l,n}$</td>
<td>LF</td>
<td>87.5 ± 2.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SIFT$_{n}$</td>
<td>LF</td>
<td>87.9 ± 2.2</td>
</tr>
<tr>
<td>FV + SVM, with NIR</td>
<td>COL$_{r,g,b}$</td>
<td>SIFT$_{l}$</td>
<td>LF</td>
<td>84.5 ± 2.3</td>
</tr>
<tr>
<td>Brown and Süssstrunk [2]</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>72.0 ± 2.9</td>
</tr>
</tbody>
</table>

# Better Features and Classifiers

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Descriptor1</th>
<th>Descriptor2</th>
<th>Fusion type</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{COL}_{r,g,b} )</td>
<td>( \text{SIFT}_{l,n} )</td>
<td>( \text{SIFT}_{n} )</td>
<td>EF</td>
<td>84.3 ± (1.7)</td>
</tr>
<tr>
<td>( \text{COL}_{r,g,b,n} )</td>
<td>( \text{SIFT}_{l,n} )</td>
<td>( \text{SIFT}_{n} )</td>
<td>EF</td>
<td>84.1 ± (2.6)</td>
</tr>
<tr>
<td>( \text{COL}_{p1,p2,p3} )</td>
<td>( \text{SIFT}_{l,n} )</td>
<td>( \text{SIFT}_{n} )</td>
<td>EF</td>
<td>83.3 ± (3.4)</td>
</tr>
<tr>
<td>( \text{COL}_{p1,p2,p3,p4} )</td>
<td>( \text{SIFT}_{l,n} )</td>
<td>( \text{SIFT}_{n} )</td>
<td>EF</td>
<td>85.8 ± (2.6)</td>
</tr>
</tbody>
</table>

## FV + SVM, with NIR

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Descriptor1</th>
<th>Descriptor2</th>
<th>Fusion type</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{COL}_{r,g,b} )</td>
<td>( \text{SIFT}_{l,n} )</td>
<td>( \text{SIFT}_{n} )</td>
<td>EF</td>
<td>87.4 ± (2.7)</td>
</tr>
<tr>
<td>( \text{COL}_{r,g,b,n} )</td>
<td>( \text{SIFT}_{l,n} )</td>
<td>( \text{SIFT}_{n} )</td>
<td>LF</td>
<td>86.5 ± (2.4)</td>
</tr>
<tr>
<td>( \text{COL}_{p1,p2,p3,p4} )</td>
<td>( \text{SIFT}_{l,n} )</td>
<td>( \text{SIFT}_{n} )</td>
<td>LF</td>
<td>87.5 ± (2.4)</td>
</tr>
</tbody>
</table>

## FV + SVM, no NIR

|         | \( \text{COL}_{r,g,b} \) | \( \text{SIFT}_{l} \) | LF | 84.5 ± (2.3) |

Brown and Süssstrunk [2]

Conclusions

• Using an extra (non-visible) NIR channel has many benefits in multimedia applications.
  - Enhance the content
  - Extract more accurate information about the scene.
Other RGB-NIR Applications

- Microsoft Kinect [www.primesense.com]

- Projector uses 830nm (NIR) laser
- NIR camera is tuned to this frequency
- Projector-camera system to capture depth (stereo)
More Physics...

- Cameras (3D)
- Microphones
- GPS
- Accelerometer
- Digital Compass
- Ambient Light
- Temperature?
- Air Quality?
- ...

Photo Courtesy Chris Jordan

Conclusions

• Using an extra (non-visible) NIR channel has many benefits in multimedia applications.  
  - Enhance the content  
  - Extract more accurate information about the scene.

• There are many sensors providing physical measurements that can improve imaging applications.  
  - And we have only begun to exploit their possibilities.
Thank you!

sabine.susstrunk@epfl.ch
http://ivrg.epfl.ch/publications/