Imperial College London

INTERNATIONAL HELLENIC UNIVERSITY

Image Fusion: Theory and Application

Dr. Nikolaos Mitianoudis

INFORMATICS & TELEMATICS INSTITUTE
Centre of Research & Technology - Hellas

28 April 2010

OUTLINE

- Introduction to Image Fusion
- Transform-domain Image Fusion
- Extraction of ICA / Topographic ICA bases
- Fusion Rules using ICA Fusion
- Contrast Correction for Multi-modal ICA fusion
- Conclusions.

WHAT IS IMAGE FUSION ?

Various optical sensors of different modality are currently available

a) Visual sensors
b) Infrared sensors
c) Gamma sensors
d) X-Ray sensors etc

Each of these modalities offer different information to the human operator or a computer vision system.

Hitherto, an intelligent system would have to examine these inputs **independently** to make an inference.

WHAT IS IMAGE FUSION ?

For example, a visual sensor can provide interesting visual, textural details

WHAT IS IMAGE FUSION ?

An infrared sensor can provide less detailed visual, textural details but can highlight the man hiding behind the smoke

WHAT IS IMAGE FUSION ?

Instead of processing the individual images, we can transfer (**fuse**) all the interesting details to a single composite image

IMAGE FUSION

Definition: Image fusion is the process of combining information from different sensors that capture the same scene.

Objective: Enhance the **perception quality of the observed scene**, not achievable by a single sensor.

Assumption: All input images are assumed to be registered, prior fusion

(a) Some Image Fusion systems operate in the **spatial domain**.

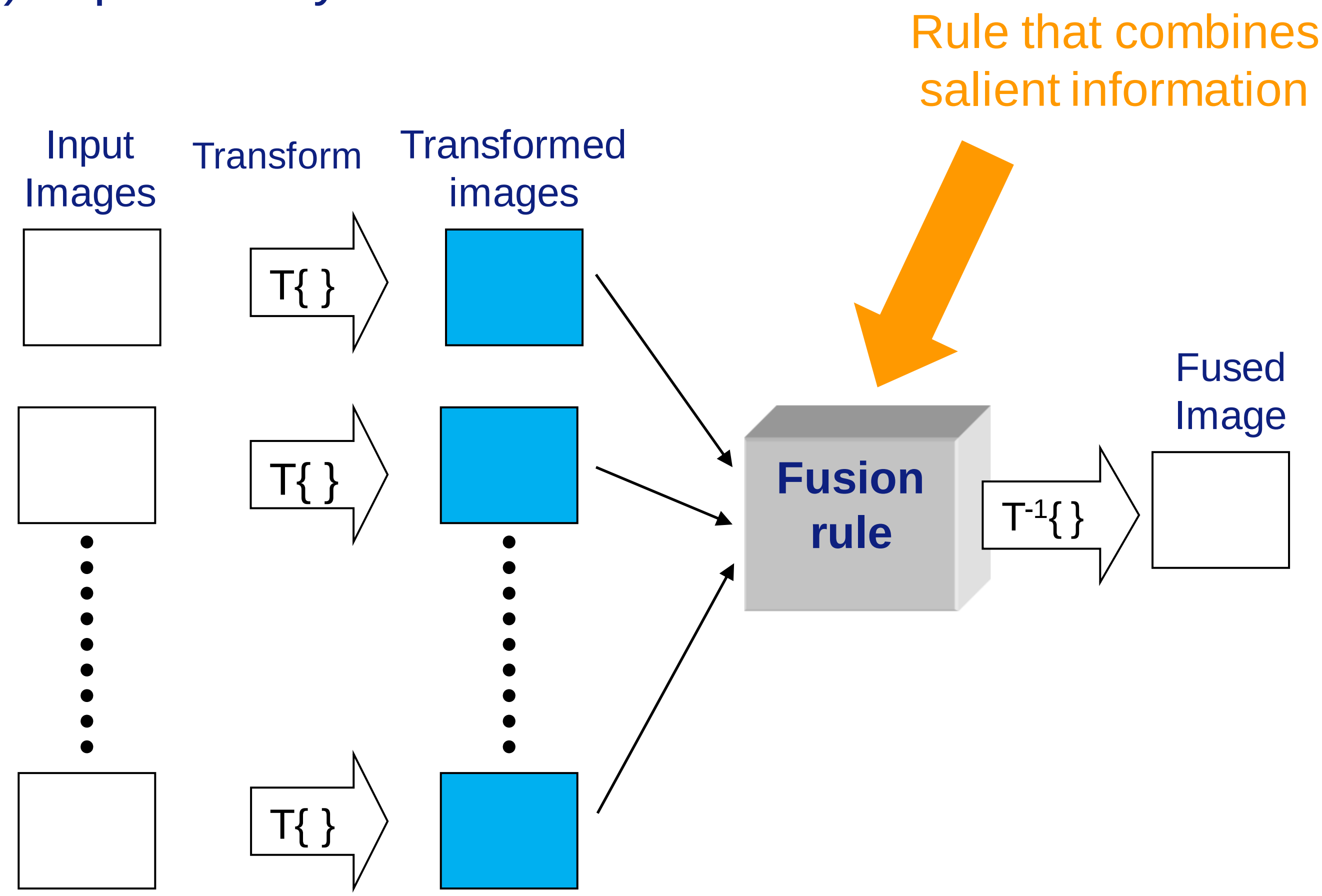
- Less computationally expensive, no transformation needed.
- Benefit from previous work on spatial-domain enhancement

(b) Most Image Fusion systems perform in the **transform domain** in order to highlight and enhance **image salient features**.

TRANSFORM-DOMAIN IMAGE FUSION

Common Transformation choices:

a) Wavelet-Decomposition
b) Dual-Tree Wavelet Decomposition (Approximate shift invariance)
c) Laplacian Pyramids



TRANSFORM-DOMAIN IMAGE FUSION

Proposal:

Replace **DTWT** with **ICA** and **Topographical ICA bases** trained on similar images.

Benefits:

(a) **Better performance**, (transform is **tailored to application**).
(b) **More degrees of freedom** than DTWT.
(c) Describe **image features more accurately**.

Drawbacks:

The bases are **shift variant**

No problem, if images are **registered**.

Sliding window to overcome problem, however computationally expensive.

WHAT ARE ICA BASES ?

(Hyvarinen et al 2001)

ICA bases offer a customised transformation by training on similar content images through Independent Component Analysis

ICA bases are essentially localised, bandpass filters with arbitrary degrees of freedom.

- They resemble Gabor filters

- They have no ordering.

- They are not shift invariant.

INSIGHTS FROM THE HUMAN VISUAL SYSTEM

Receptive fields of simple cells in mammalian primary visual cortex can be characterized as spatially localized, oriented and bandpass.

(Field 1996, Olshausen-Field @ Nature 1996)

Such filter responses should emerge from:

a) Unsupervised learning algorithms that estimate a factorial product of independent visual features. (Barlow 1989, Bell 1996)

b) Unsupervised learning algorithms that attempt to find sparse linear codes for natural scenes.

(Olshausen-Field 1996, Lewicki-Sejnowski 2000, Hyvarinen-Hoyer-Oja 2001)

INSIGHTS FROM THE HUMAN VISUAL SYSTEM

The image $I(x, y)$ can be expressed as a linear combination of basis functions $b_i(x, y)$:

$$I(x, y) = \sum_i u_i b_i(x, y)$$

where u_i are scaling coefficients, such that $u_i = \langle I(x, y), b_i(x, y) \rangle$.

The requested visual fields can be estimated either by :

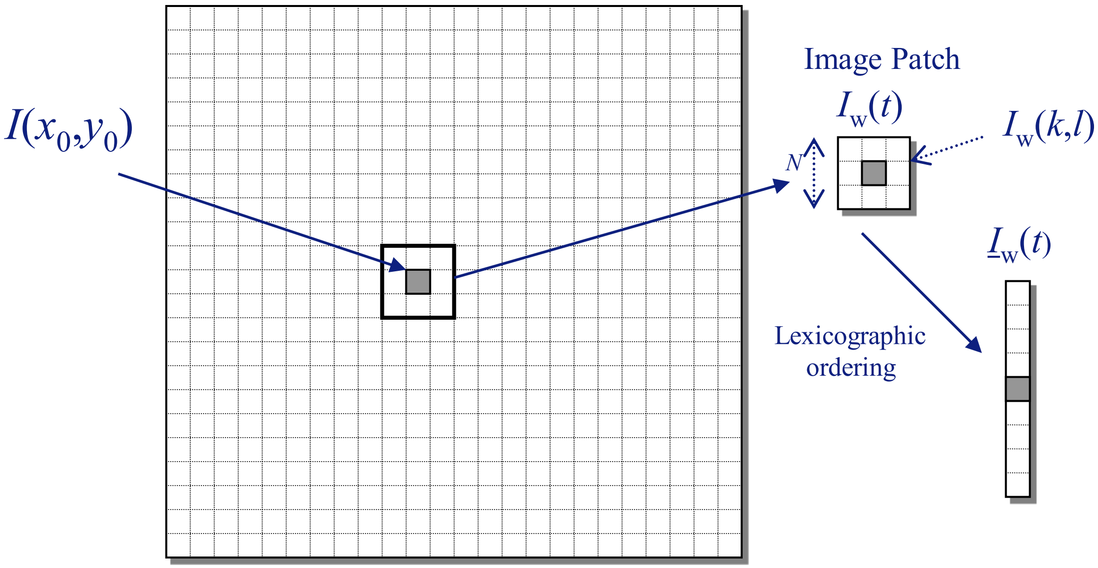
- Maximising the sparseness of u_i .
- Maximising the statistical independence of $b_i(x, y)$.

Independent Component Analysis (ICA) can perform these two equivalent tasks, i.e. estimate analysis bases, resembling receptive fields of the visual cortex.

(Lewicki-Sejnowski 2000, Hyvarinen-Hoyer-Oja 2001)

TRAINING LOCAL IMAGE ANALYSIS BASES

Assume image $I(x,y)$ of size $M_1 \times M_2$ and window W of size $N x N$ centered around pixel (x_0, y_0).



Express the observed patches:

Transform coefficients

$$\underline{u}(t) = A\underline{I}_W(t)$$

$$\underline{I}_W(t) = B\underline{u}(t) = A^{-1}\underline{u}(t)$$

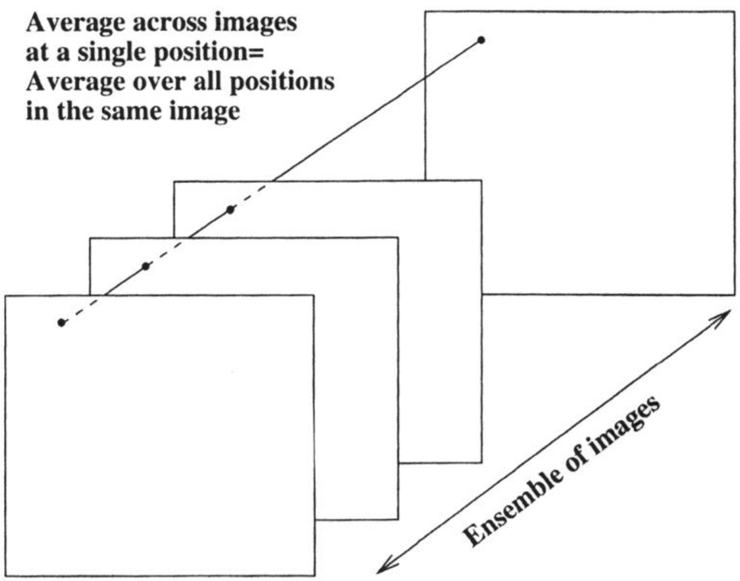
(A analysis kernel)

(B synthesis kernel)

ERGODICITY IN LOCAL IMAGE ANALYSIS

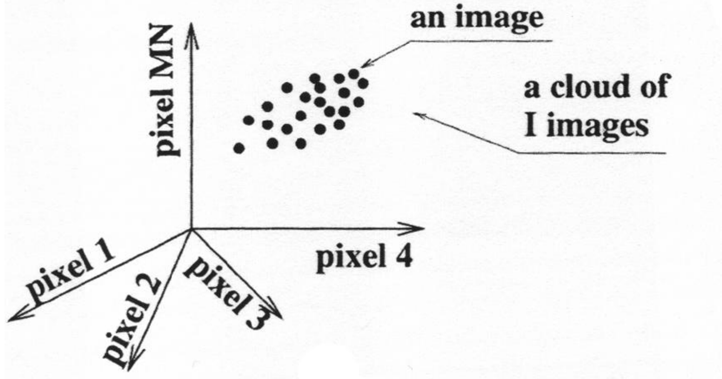
(M. Petrou and C. Petrou 2010)

The population of extracted patches assumes ergodicity, i.e.



Natural variability of images = natural variability of each image.

In our statistical estimation experiments, each extracted patch represents a dot in the next diagram.



TRAINING PRINCIPAL COMPONENT ANALYSIS BASES

Estimate the analysis kernel $A = [\underline{a}_1 \quad \underline{a}_2 \quad .... \quad \underline{a}_N]^T$.

Principal Component Analysis (PCA) bases:

PCA can identify **uncorrelated vector bases**, given linear generative model.

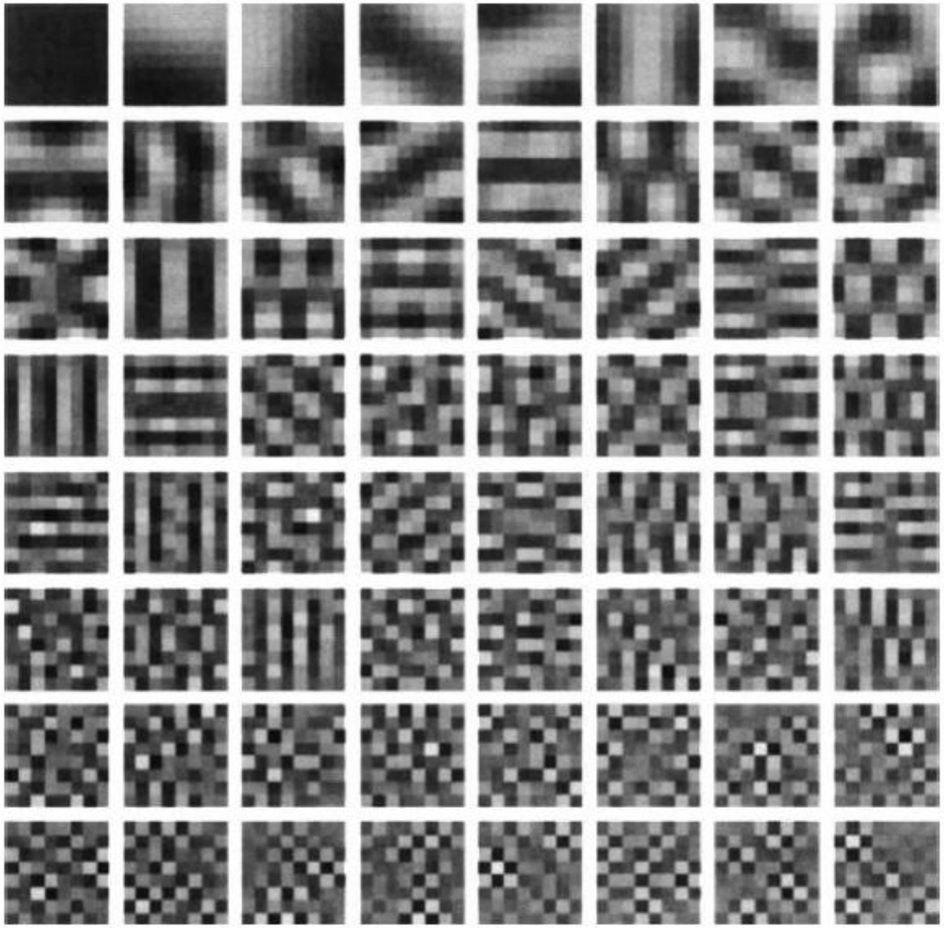
Assuming **uncorrelatedness**, the analysis bases $[\underline{a}_1 \quad \underline{a}_2 \quad .... \quad \underline{a}_N]^T$ are given by **eigenvectors** of **data correlation matrix** $C = E\{\underline{I}_W \underline{I}_W^T\}$ normalised by corresponding eigenvalues.

Dimensionality reduction is also possible, forming a $K \times N^2$ reduced orthonormal analysis kernel.

PCA bases, using **second-order statistics**, estimate **Gaussian-like localised bases**.

TRAINING PRINCIPAL COMPONENT ANALYSIS BASES

PCA bases trained using local 8×8 patches from natural images.



PCA bases do not estimate localised structures.

TRAINING OF ICA AND TOPOGRAPHIC ICA BASES

Independent Component Analysis (ICA) bases:

A stricter criterion is **statistical independence** between the transform coefficients.

Assuming **statistical independence**, the estimated analysis bases are **localised edge features**, resembling Gabor functions.

Independent Component Analysis can estimate the synthesis matrix B of the following linear generative model:

$$\underline{I}_W(t) = B\underline{u}(t)$$

Conditions:

- B is a full rank matrix, i.e. redundancy has been removed by PCA. ($\mathrm{rank}(B) = K$)

TRAINING OF ICA AND TOPOGRAPHIC ICA BASES

An interpretation of sparsity or statistical independence is nonGaussianity.

Enhance sparsity by maximizing nonGaussianity => **FastICA** algorithm (Hyvarinen 1997)

Training using **FastICA** algorithm. PCA bases estimation is required. Reduced K bases set also possible $Q = \left[\underline{q}_1 \quad \underline{q}_2 \quad .... \quad \underline{q}_K\right]^T$.

Estimation via iterating over a training set of patches:

$$\underline{q}_i^+ \leftarrow E\{\underline{q}_i \phi(\underline{q}_i^T \underline{u}_{PCA})\} - E\{\phi'(\underline{q}_i^T \underline{u}_{PCA})\}\underline{q}_i \qquad \forall \quad i = 1, \ldots, K$$

$$Q \leftarrow Q(Q^T Q)^{-0.5}$$

where $\phi(y) = -\partial G(y) / \partial y$ and $G(y)$ any non-quadratic function.

TRAINING OF ICA AND TOPOGRAPHIC ICA BASES

Topographic Independent Component Analysis (TopoICA) bases:

Statistical independence is a strong assumption.

Topographic ICA allows **spatial correlation** between neighbouring bases. PCA is again a prerequisite.

Estimation via iterating over a training set of patches:

$$\underline{q}_i^+ \leftarrow \underline{q}_i + \eta E\{\underline{u}_{PCA}(\underline{q}_i^T \underline{u}_{PCA}) r_i\} \quad \forall \quad i = 1, \ldots, K$$

$$r_i \leftarrow \sum_{k=1}^{K} h(i,j)\phi\Big(\sum_{j=1} h(j,k)(\underline{q}_i^T \underline{u}_{PCA})^2\Big) \qquad h(i,j) = \begin{cases} 1, & \text{if } |i-k| \le L \\ 0, & \text{otherwise} \end{cases}$$

$$Q \leftarrow Q(Q^T Q)^{-0.5}$$

where η is the learning rate, $\phi(y) = -\partial G(y)/\partial y$ and L is the neighbourhood size.

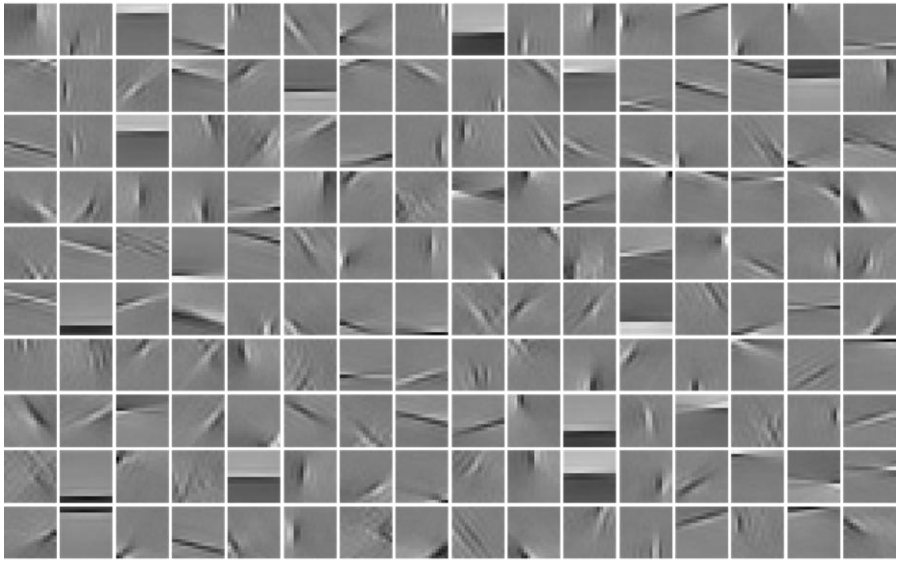
TRAINING OF ICA AND TOPOGRAPHIC ICA BASES

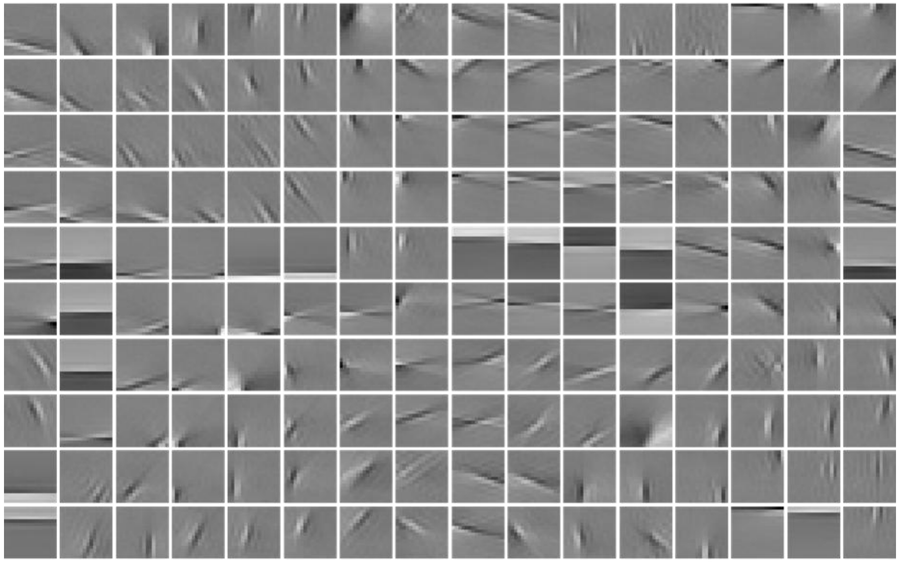
The ICA and Topographic ICA bases are given by the product of respective ICA matrix Q and PCA bases A_{PCA}.

$$A \leftarrow QA_{PCA}$$

ICA AND TOPOGRAPHICAL BASES

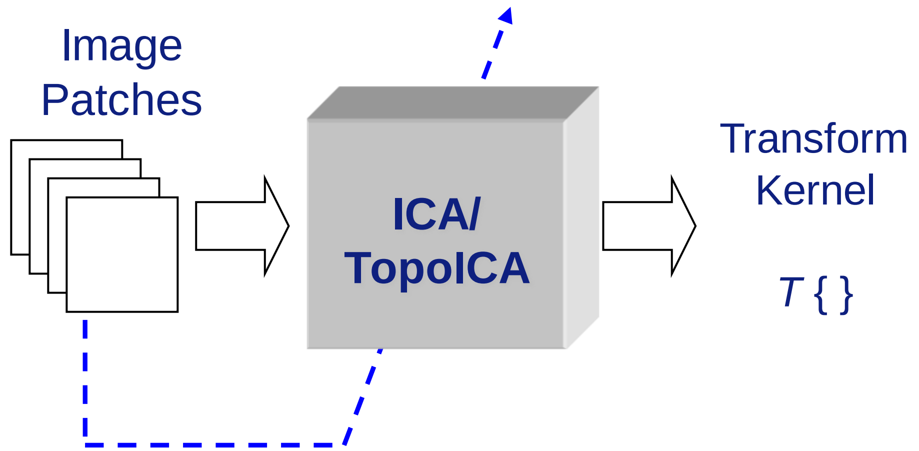
Trained ICA and Topographical ICA bases using $16x16$ patches from natural images.





Difference: Topography, i.e. allowed spatial correlation.

TRAINING OF ICA AND TOPOGRAPHIC ICA BASES



Training strategy:

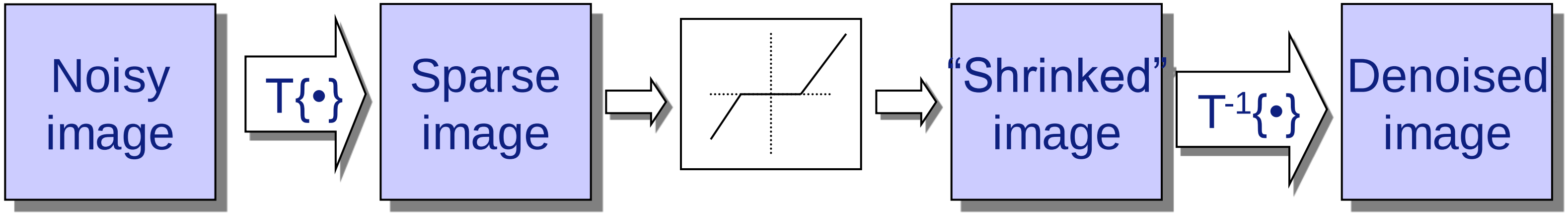
- Randomly select ~10000 8×8 patches from similar content images.
- Select 40-64 bases using PCA.
- Perform ICA or TopoICA to train corresponding bases.
- Store the analysis kernel.

DENOISING VIA SPARSE CODE SHRINKAGE

(Hyvarinen-Hoyer-Oja 2001)

Concept:

- ➢ Apply a sparse linear transform T{•} on the noisy image.
- ➢ In the sparse representation perform ML estimation using superGaussian priors (sparse code shrinkage).
- ➢ Move the "shrinked" data back to the original domain.



ML estimator assuming isotropic noise $C_N = \sigma^2 I$:

$$S(x,y) = A^{-1} g(AI(x,y))$$

Noisy Image

Denoised Image

Shrinkage model

DENOISING VIA SPARSE CODE SHRINKAGE

(Hyvarinen-Hoyer-Oja 2001)

ICA
bases

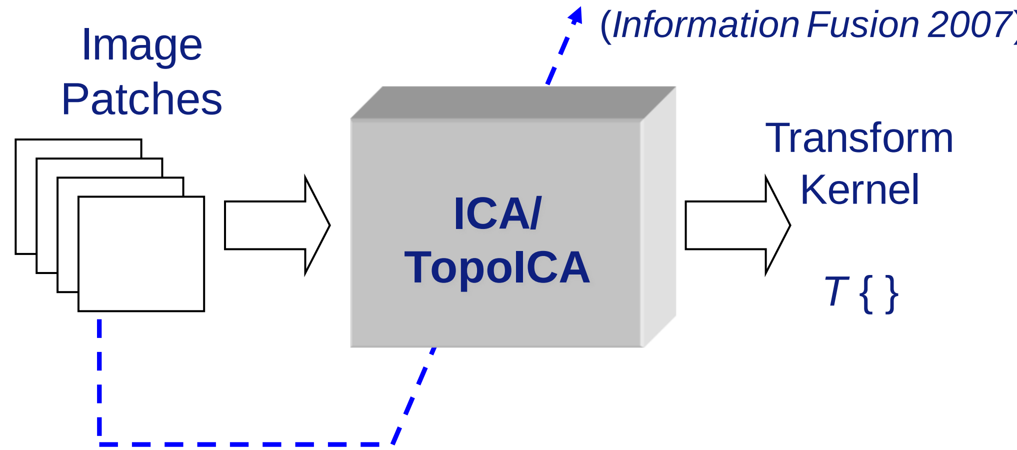
Topographic
ICA
bases

SNR=20.4156 dB

SNR=21.1336 dB

IMAGE FUSION USING ICA AND TOPOGRAPHIC ICA

Transform estimation



Fusion

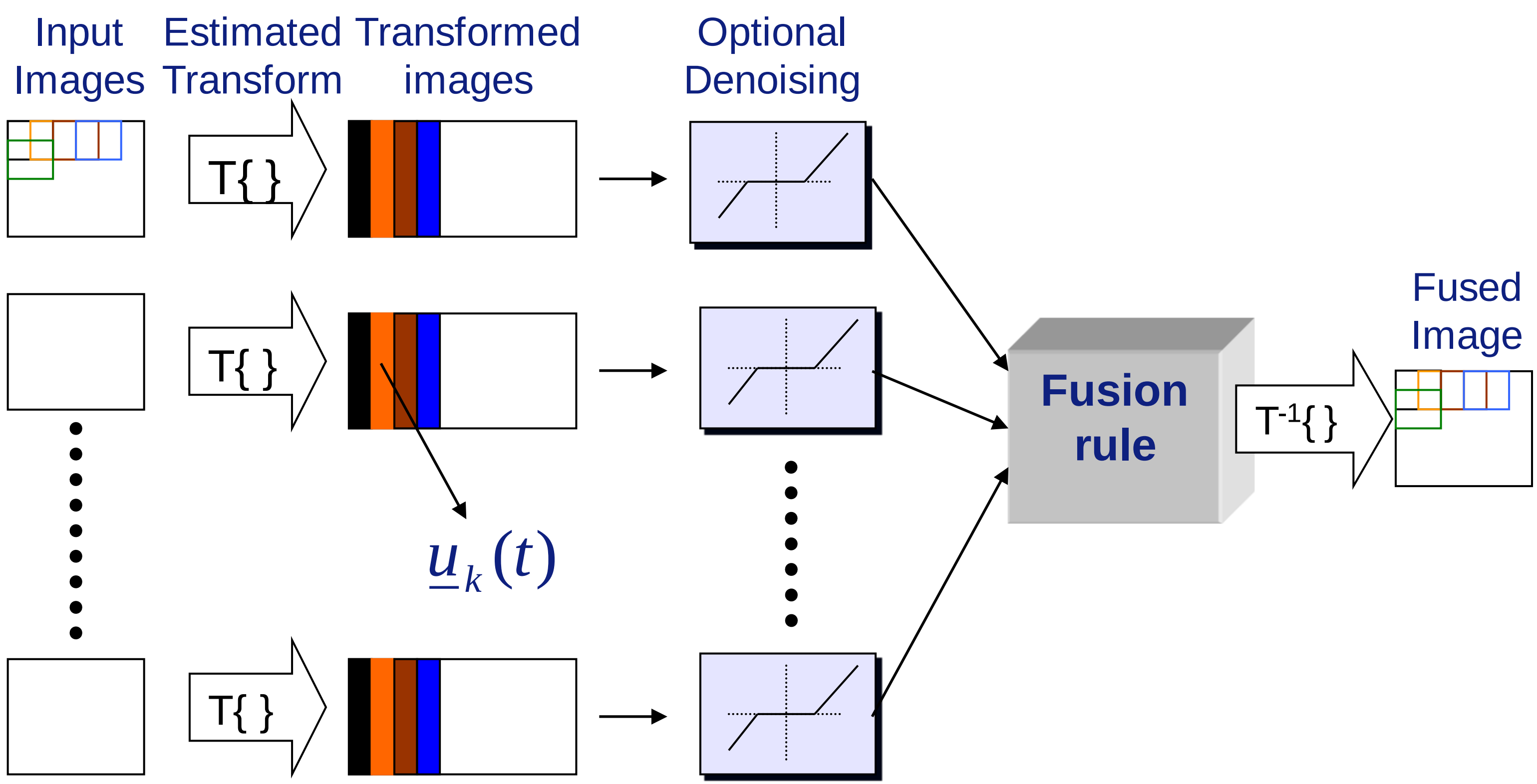
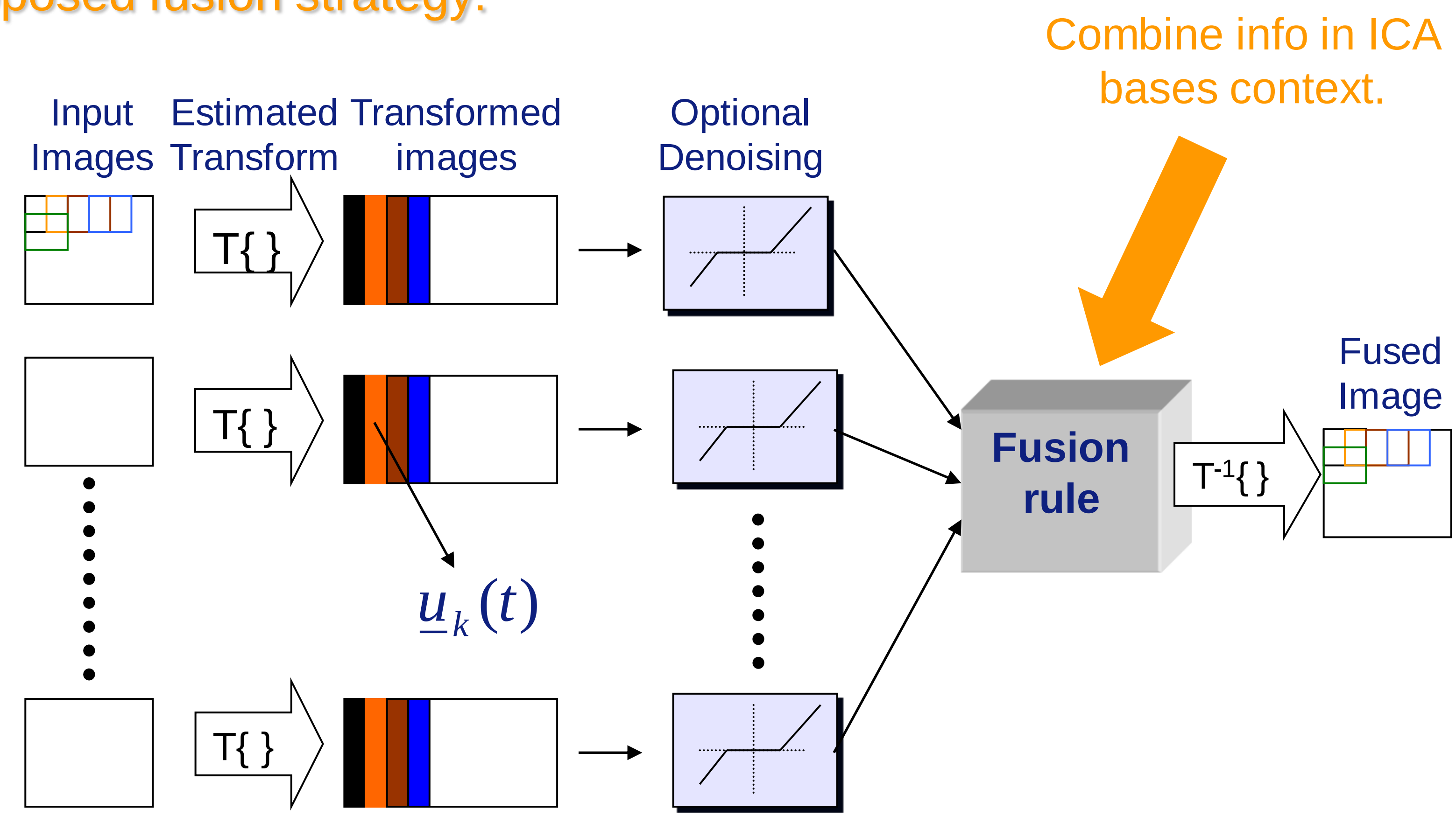


IMAGE FUSION USING ICA BASES

(*Information Fusion 2007*)

Proposed fusion strategy:



8×8 patches are selected using 1-pixel overlap from input images.

GENERAL FUSION RULES

(*Information Fusion 2007*)

Fusion rule = method to combine coefficient to form fused image.

Fusion by "max-abs"

Select the maximum in absolute value coefficient from all inputs.

$$T\{\underline{I}_f(t)\} = \operatorname{sgn}(T\{\underline{I}_k(t)\}) \max_k |T\{\underline{I}_k(t)\}|$$

Pros

Enhances contribution of certain bases, i.e. localised features, therefore enhances edges.

Cons

Distorts constant background or light textural information.

GENERAL FUSION RULES

(*Information Fusion 2007*)

Fusion by “mean”

Select the mean of the corresponding input coefficients.

$$T\{\underline{I}_f(t)\} = \operatorname*{mean}_k (T\{\underline{I}_k(t)\})$$

Pros

Preserves constant background information.

Cons

Smoothes out edges, as averaging is a low-pass filtering process.

Objective

Find fusion rules that tackle the shortcomings of these approaches.

WEIGHTED - COMBINATION FUSION RULE

(*Information Fusion 2007*)

Combine the image patches $\underline{I}_k(t)$ using an **activity detector** (L_1-norm):

$$E_k(t) = \| \underline{u}_k(t) \|_1 \qquad k = 1,..,T$$

The “**fused**” image is constructed, as follows:

$$T\{\underline{I}_f(t)\} = \sum_{k=1}^{T} w_k(t) T\{\underline{I}_k(t)\}$$

where

$$w_k(t) = E_k(t) / \sum_{k=1}^{T} E_k(t)$$

As the ICA bases capture activity $\longrightarrow$ large $E_k(t)$ represents larger activity.

The weighted combination emphasises patches with greater activity.

REGION-BASED FUSION RULE

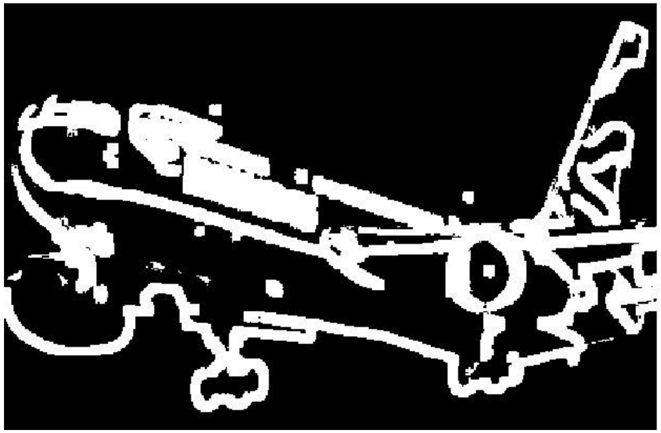
(*Information Fusion 2007*)

Divide the image using activity detector $E_k(t)$ into two regions:

a) **active** regions (**salient features**)
if $E_k(t) \geq 2\ \mathrm{mean}_t\{E_k(t)\}$

b) **non-active** regions (**constant background**)
if $E_k(t) < 2\ \mathrm{mean}_t\{E_k(t)\}$

Segmentation map in two regions



Fuse
- **active** regions with the **weighted combination** rule
- **non-active** regions with the **mean** rule.

ADAPTIVE FUSION RULES

(*Information Fusion 2007*)

Objective: Identify a self-adaptive fusion scheme emphasising local features in most fusion scenarios.

Formulation: $$\underline{u}_f = w_1 \underline{u}_1 + w_2 \underline{u}_2 + \ldots + w_T \underline{u}_T$$

Fused image patch

Fusion weights

Input image patches (lexicographic ordering)

Problem: Estimate $\underline{w} = \begin{bmatrix} w_1 & w_2 & \ldots & w_T \end{bmatrix}^T$, so that local features are emphasised.

Assuming $N \times N$ patches, we can define

$$\underline{x}(n) = \begin{bmatrix} u_1(n) & u_2(n) & \ldots & u_T(n) \end{bmatrix}^T, \quad \forall \quad n = 1, \ldots, N^2$$

The fused image can be expressed, as follows:

$$u_f(n) = \underline{w}^T \underline{x}(n), \quad \forall \quad n = 1, \ldots, N^2$$

ADAPTIVE FUSION RULES

(ICASSP 2006)

Observation:

The actual non-distorted representation should be ***more sparse*** *than the distorted or different sensor inputs*.

Conclusion:

The fusion process should **enhance sparsity** in the ICA domain.

Approach:

Perform fusion using **Maximum Likelihood** (ML) estimation of $J(\underline{w})$ subject to several constraints:

$$\max_{\underline{w}} J(\underline{w}) = \max_{\underline{w}} E\{\log p(u_f)\}$$

$$\text{subject to} \quad [1 \quad 1 \quad \dots \quad 1]\underline{w} = \underline{e}^T \underline{w} = 1$$

$$\underline{w} \geq 0$$

ADAPTIVE FUSION RULES

(ICASSP 2006)

Sparse models:

Model **sparsity**

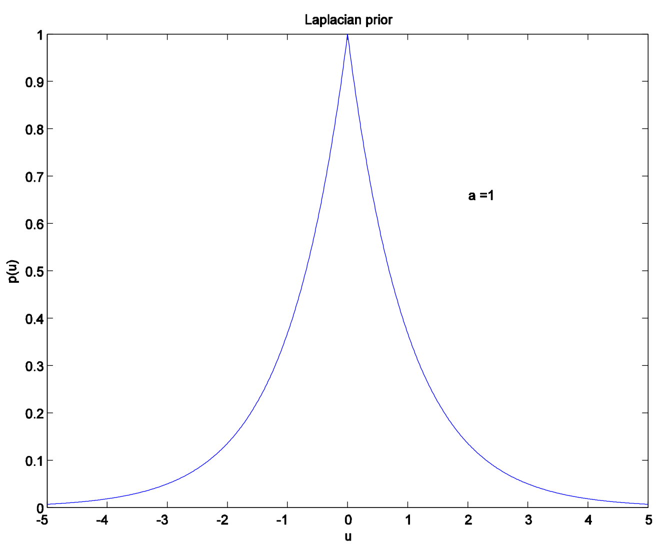
α, *s* const

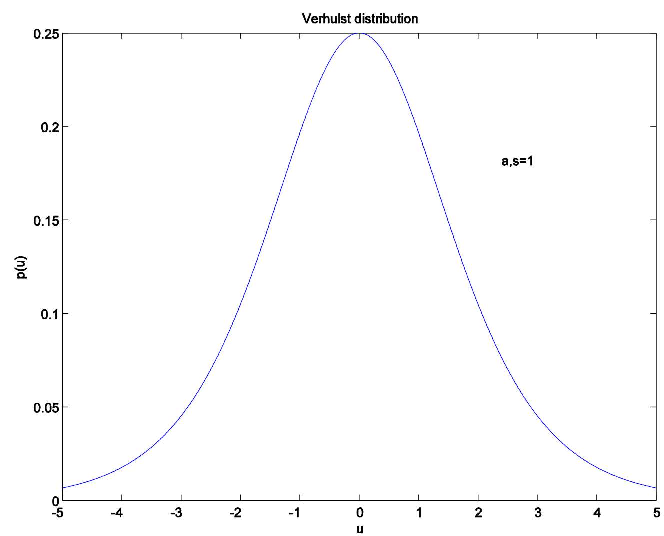
Laplacian distribution

$$p(u) \propto e^{-a|u|}$$

Verhulstian distribution

$$p(u) \propto \frac{e^{-u/s}}{s(1+e^{-u/s})^2}$$





VARIOUS FUSION RULES FOR ICA-BASED FUSION

(*ICASSP 2006, Information Fusion 2007*)

- "Max-Abs" rule

  The greatest coefficients in abs construct the fused image.

- "Mean" rule

  An average of the input coefficients constructs the fused image.

- Weighted-Combination rule:

  Input coefficients are weighted according to contribution of the total energy.

- Regional Rule:

  Segment scene into active/non-active areas and use different rules for each area.

- Adaptive Rule:

  Estimate optimal weights for the coefficients by ML-estimation for a sparse prior (e.g. Laplacian prior).

EXAMPLE

Fusion of <u>out-of-focus</u> images:

Input Image 1

Input Image 2

DT-WT (max-abs)

TopoICA (max-abs)

TopoICA (WC)

TopoICA (region)

TopoICA (Laplace)

TopoICA (Verhulst)

Tested Transforms Dual-Tree Wavelet Transform (DTWT), ICA, Topographic ICA.

Tested fusion schemes Max-abs, Weighted Combination (WC), Region-based (region), Adaptive (Laplacian).

EXAMPLE

Fusion of out-of-focus images:

Performance comparison of several combinations of transforms and fusion rules for out-of-focus datasets, in terms of the Piella/Petrovic indexes.

	WP (Sym7)	DT-WT	ICA	TopoICA
		Clocks dataset		
Max-abs	0.8727/0.6080	0.8910/0.6445	0.8876/0.6530	0.8916/0.6505
Mean	0.8747/0.5782	0.8747/0.5782	0.8523/0.5583	0.8560/0.5615
Weighted	–	–	0.8678/0.6339	0.8743/0.6347
Regional	–	–	0.8583/0.5995	0.8662/0.5954
Laplacian	–	–	0.8521/0.5598	0.8563/0.5624
		Disk dataset		
Max-abs	0.8850/0.6069	0.8881/0.6284	0.9109/0.6521	0.9111/0.6477
Mean	0.8661/0.5500	0.8661/0.5500	0.8639/0.5470	0.8639/0.5459
Weighted	–	–	0.9134/0.6426	0.9134/0.6381
Regional	–	–	0.9069/0.6105	0.9084/0.6068
Laplacian	–	–	0.8679/0.5541	0.8655/0.5489

EXAMPLE

Fusion of multi-modal images:

HS IR input

MicroLW input

CCD input

DT-WT (max-abs)

TopoICA (max-abs)

TopoICA (WC)

TopoICA (Laplace)

TopoICA (Verhulst)

Tested Transforms Dual-Tree Wavelet Transform (DTWT), ICA, Topographic ICA.

Tested fusion schemes Max-abs, Weighted Combination (WC), Region-based (region), Adaptive (Laplacian).

EXAMPLE

Fusion of surveillance images:

Visual Sensor

Infrared sensor

DT-WT (max-abs)

TopoICA (max-abs)

TopoICA (WC)

TopoICA (Region)

TopoICA (Laplace)

TopoICA (Verhulst)

Tested Transforms Dual-Tree Wavelet Transform (DTWT), Topographic ICA.
Tested fusion schemes Max-abs, Weighted Combination (WC), Region-based (region), Adaptive (Laplacian - Verhulst).

OPTIMAL CONTRAST FOR MULTI-MODAL ICA FUSION

(IEEE Sensors Journal 2008)

Fact

The ICA-based fusion performs fusion of local-patches from the different sensor images.
These patches need to be normalised to zero mean.

Question

How do we reconstruct the means of the fused image ?

Previous Solution

The means were reconstructed using an average of the input sensor means.

Problems

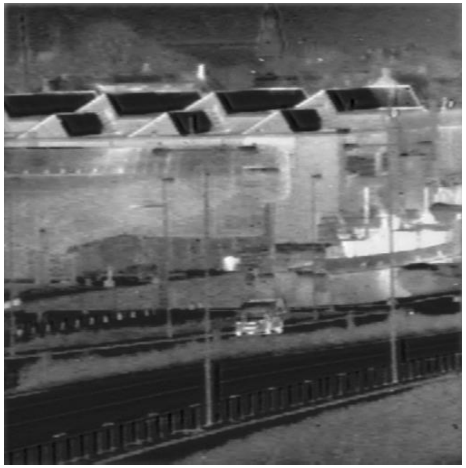
- Correct solution in the case of out-of-focus examples.
- Might not be optimal in the case of multi-modal examples.

EXAMPLE OF MEANS CHOICE

(IEEE Sensors Journal 2008)

Visual Sensor

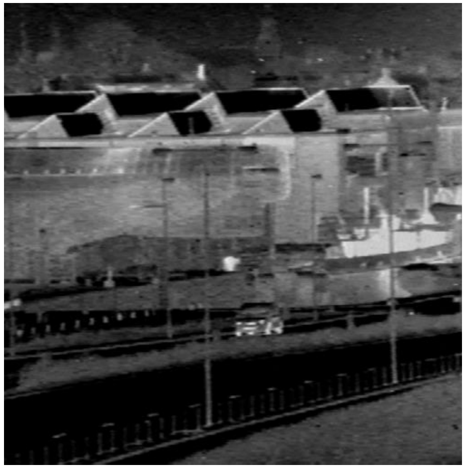
IR Sensor



Means from Visual

Average Means

Means from IR



A GENERALISED OPTIMAL CONTRAST SCHEME

(IEEE Sensors Journal 2008)

A general cost function seeking to identify the weights of the means. was formulated, based on a simplified Piella index.

$$m_f = w_1 m_{x_1} + w_2 m_{x_2} + \ldots + w_T m_{x_T}$$

Fused image mean

weights

Input means

Piella's index is based on the Wang and Bovik index:

$$Q(x,f) = \frac{2\sigma_{xf}}{\sigma_x^2 + \sigma_f^2} \frac{2m_x m_f}{m_x^2 + m_f^2}$$

Wang and Bovik index is divided into a variance and means term:

$$Q(x,f) = Q_\sigma(x,f) Q_m(x,f) = \underbrace{Q_\sigma(x,f)}_{\text{constant}} \frac{2m_x m_f}{m_x^2 + m_f^2}$$

A GENERALISED OPTIMAL CONTRAST SCHEME

(IEEE Sensors Journal 2008)

Optimise the Piella Index in terms of $\underline{w} = [w_1 \quad w_2 \quad \dots \quad w_N]^T$.

Gradient ascent on the cost function will give the following iterative update for the weights:

$$\underline{w}^{+} \leftarrow \underline{w} + \eta \mathcal{E} \left\{ \underline{m}_x \sum_{i=1}^{T} A_{\sigma_{x_i}} m_{x_i} \frac{m_{x_i}^2 - m_f^2}{(m_{x_i}^2 + m_f^2)^2} \right\}$$

where $A_{\sigma_{x_i}} = \lambda_i Q_\sigma(x_i, f)$

To avoid unnecessary deviations, during the updates we can force the weights to sum up to 1.

$$\underline{w}_i^{+} \leftarrow \underline{w} / ([\, 1 \quad 1 \quad \dots \quad 1 \,]\underline{w})$$

A GENERALISED OPTIMAL CONTRAST SCHEME

(IEEE Sensors Journal 2008)

(a) InfraRed Image

(b) Visual Image

(c) Equal Weights ($Q = 0.7920$)

(d) Optimised Weights ($Q = 0.7968$)

(e) DT-WT ($Q = 0.7758$)

A GENERALISED OPTIMAL CONTRAST SCHEME

(*IEEE Sensors Journal 2008)*

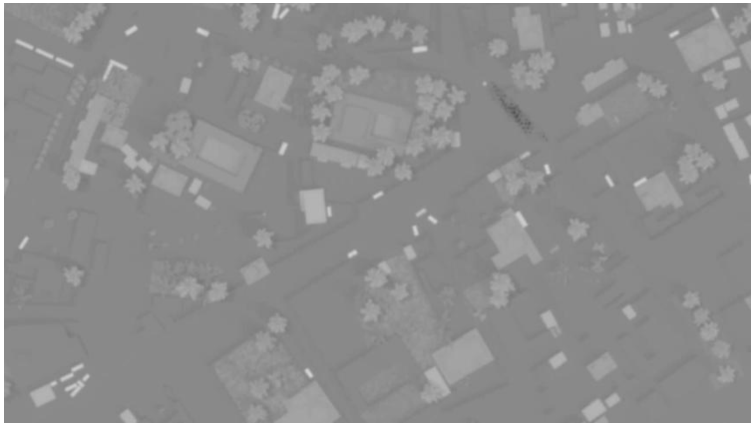
Method	''Dune''	''Trees''	''Uncamp''	Octet 1	Octet 2	Car Image
ICA - Equal Weights	0.7311	0.7770	0.7441	0.8251	0.8176	0.6822
ICA - Opt. Weights	0.7325	0.7814	0.7452	0.8354	0.8677	0.6857
DT-WT	0.7156	0.7595	0.7317	0.8254	0.8602	0.6392

VIDEO FUSION EXAMPLES

(AMDF Project 2003-2009)

RGB Camera

HS Camera



Fusion Result

IMPROVED TARGET FOLLOWING WITH FUSION

(AMDF Project 2003-2009)

Tracking on RGB video

Tracking on Fused video

VIDEO FUSION EXAMPLES

(AMDF Project 2003-2009)

RGB Camera

LWIR Camera

Fusion Result

IMPROVED TARGET FOLLOWING WITH FUSION

(AMDF Project 2003-2009)

Tracking on RGB video

Tracking on LWIR video

Tracking on Fused video

CONCLUSIONS

- Proposed an alternative image fusion framework, based on trained ICA bases.

- Improved results compared to Dual-Tree Wavelets:
  - Transform is tailored to application.
  - More degrees of freedom than DTWT.

- Proposed different fusion rules and techniques for out-of-focus and multi-modal fusion.

- Fusion improves target tracking performance

REFERENCES

[1] T. Stathaki, *Image fusion: Algorithms and Applications*, Academic Press, pages 520, 2008.

[2] M. Petrou, C. Petrou, *Image Processing: The Fundamentals*, Wiley, pages , 2010.

[3] N. Kingsbury. *The dual-tree complex wavelet transform: a new technique for shift invariance and directional filters*, Proc. IEE Digital Signal Processing Workshop, volume 86, 1998.

[4] G. Piella. *A general framework for multiresolution image fusion: from pixels to regions*. Information Fusion, 4:259--280, 2003.

[5] A. Hyvarinen, P.O. Hoyer, and M. Inki, *Topographic independent component analysis*, Neural Computation, 13, 2001.

[6] J.J. Lewis, R.J. O'Callaghan, S.G. Nikolov, D.R. Bull, and C.N. Canagarajah. *Region-based image fusion using complex wavelets*, In Proc. 7th International Conference on Information Fusion, pages 555--562, Stockholm, Sweden, 2004.

[7] N. Mitianoudis, T. Stathaki, *Pixel-based and Region-based Image Fusion schemes using ICA bases*, Information Fusion 8 (2), pp. 131–142, April 2007.

[8] N. Mitianoudis, T. Stathaki, *Optimal Contrast Correction for ICA-based Fusion of Multimodal Images*, IEEE Sensors Journal, Vol. 8, No. 12, pp. 2016 - 2026, Dec. 2008.

ACKNOWLEDGEMENTS

This work was supported by the Applied Multi-Dimensional Fusion (AMDF) project of the Defence Technology Centre, UK.

Thanks for your attention !