Image Redundancy and Non-Parametric Estimation for Image Representation

Charles Kervrann, Patrick Pérez, Jérôme Boulanger

INRIA Rennes / INRA MIA Jouy-en-Josas / Institut Curie Paris
VISTA Team
http://www.irisa.fr/vista/
Campus universitaire de Beaulieu
35042 Rennes cedex - France

Change detection problem

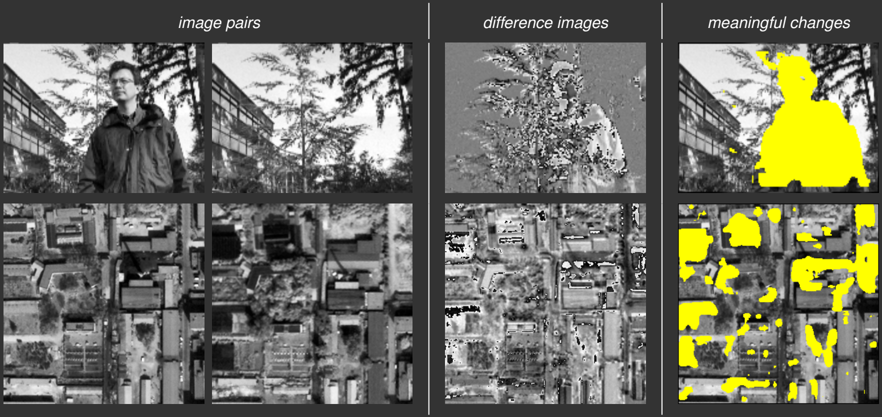


FIG.: Change detection masks superimposed on the first images, correctly highlighting the person that disappeared in the second image (top) and the buildings that appeared or disappeared in the second satellite image (bottom).

Related methods

- **Recent surveys on change detection :**

1. Radke, R., Andra, S., Al Kofahi, O., Roysam, B. : Image change detection algorithms : a systematic survey. IEEE Trans. Image Processing **14** (2005) 294–307
2. Rosin, P. : Thresholding for change detection. Comp. Vis. Image Understanding **86** (2002) 79–95
3. Aach, T., Dumbgen, L., Mester, R., Toth, D. : Bayesian illumination-invariant motion detection. In : ICIP (3), Thessaloniki, Greece (2001) 640–643

- **Graph-cuts and Markov models :**

4. Xiao, J., Shah, M. : Motion layer extraction in the presence of occlusion using graph cuts. IEEE Trans. Pattern Anal. Mach. Intell. **27** (2005) 1644–1659

- **Patch-based modeling :**

5. Leung, T.K., Malik, J. : Detecting, localizing and grouping repeated scene elements from an image. In : ECCV (1), Cambridge, UK (1996) 546–555
6. Buades, A., Coll, B., Morel, J. : Nonlocal image and movie denoising. Int. J. Comp. Vision **76** (2008) 123–139
7. Shechtman, E., Irani, M. : Matching local self-similarities across images and videos. In : CVPR, Minneapolis, Minnesota (2007) 1–8

- **A contrario modeling :**

8. Lisani, J.L., Morel, J.M. : Detection of major changes in satellite images. In : ICIP (1), Barcelona, Spain (2003) 941–944

A. Buades, B. Coll, J.M Morel, "A review of image denoising algorithms, with a new one" , SIAM Multiscale Modeling and Simulation, 4(2) (2005) 490-530

NL-mean filter
Barbara / PSNR = 30.27 db

Bayesian NL-means filter
Barbara / PSNR = 31.03 db

Bayesian Non-Local means filter : (Kervrann et al., SSVM'07)

$$ANL_{\sigma,n}\underline{u}(x) = \frac{\sum_{x_i \in D(x)} \exp -\frac{1}{2}\left(\frac{\|\underline{u}(x) - \underline{u}(x_i)\|}{\sigma} - \sqrt{2n-1}\right)^2 \underline{u}(x_i)}{\sum_{x_i \in D(x)} \exp -\frac{1}{2}\left(\frac{\|\underline{u}(x) - \underline{u}(x_i)\|}{\sigma} - \sqrt{2n-1}\right)^2}$$

Image redundancy in an image pair

- **Image model :** Let $u = (u(x))_{x \in \Omega}$ and $v = (v(x))_{x \in \Omega}$ be an image pair defined at pixel $x \in \Omega$ as :

$$\begin{aligned} u(x) &= u_0(x) + \epsilon(x), \\ v(x) &= v_0(x) + \eta(x), \end{aligned}$$

  where u_0 and v_0 are the “true” images and the “errors” ϵ and η are i.i.d. Gaussian zero-mean random variables with unknown variance σ^2.

- **Hypothesis for change detection :** Our idea is to guess a n-dimensional square patch $\underline{u}(x)$ in u from square patches $\underline{v}(x_i)$ taken in the fixed size semi-local neighborhood $\Delta(x)$ observed at point x_i in the second image v :

$$\underline{u}(x) \equiv \underline{v}(x_i), x_i \in \Delta(x) \text{ if no scene change occurs.}$$

Local score/decision mechanism for change detection (1)

- **Step 1 :** Each pixel $x_i \in \Delta(x)$ in v computes a score $z(x_i)$

$$z(x_i) \triangleq \frac{\|\underline{u}(x) - \underline{v}(x_i)\|^2}{2\sigma^2},$$

and makes a binary decision $d(x_i) = \mathbf{1}(z(x_i) \geq \tau(x))$ w.r.t. a threshold $\tau(x)$.

- **Step 2 :** Each pixel x reaches a decision $D(x) \in \{0, 1\}$:

$$D(x) = \mathbf{1}\left(\text{count}(x) \triangleq \sum_{x_i \in \Delta(x)} d(x_i) \underset{H_0}{\overset{H_1}{\gtrless}} T\right)$$

where T is a threshold related to a probability of false alarm.

Local score/decision mechanism for change detection (2)

- **Probability of change detection :**

$$\mathbb{P}\{\text{count}(x) \geq T; n | H_0\} \triangleq \sum_{k=T}^{N} \left(\begin{array}{c} N \\ k \end{array} \right) p_{i,n}^{k} (1 - p_{i,n})^{N-k}$$

is the right tail of the binomial distribution where $p_{i,n}$ denotes $\mathbb{P}\{z(x_i) \geq \tau(x); n | H_0\}$ and $N = |\Delta(x)|$.

Parametric and Gaussian modeling (1)

- **Hypothesis** H_0 **:** $z(x_i)$ follows a central chi-squared distribution, i.e. $z(x_i) \sim \chi^2_n$.
- **Hypothesis** H_1 **:** if $\underline{u_0}(x) \not\equiv \underline{v_0}(x_i), x_i \in \Delta(x)$, $z(x_i)$ is distributed according to the non-central chi-squared distribution $\chi^2_{n,\lambda}$ with unknown parameter $\lambda = \|\underline{u_0}(x) - \underline{v_0}(x_i)\|^2/2\sigma^2$.

PDF of scores

$$f(z) = \pi_0 f_0(z) + \pi_1 f_1(z)$$

where π_0 (resp. $\pi_1 = 1 - \pi_0$) is the proportion of the true null (resp. true alternative) hypothesis H_0 (resp. H_1), f_0 and f_1 denote the PDFs (central and non-central Chi-square distributions) of scores under H_0 and H_1.

Parametric and Gaussian modeling (2)

- **Inference of a unique threshold τ for patch recognition :**

$$\mathbb{P}\{z \geq \tau\} \;=\; \pi_0 \int_\tau^\infty f_0(z)dz + \pi_1 \int_\tau^\infty f_1(z)dz$$

can be computed using an EM procedure (mixture parameters) provided λ is partially know.

Limitations of Gaussian and parametric models ?

- Cons :
  1. PDF learning and EM algorithm are necessary ...but for which patch and neighborhood sizes ?
  2. What happens if the missing object size(s) is(are) large ?
  3. Are PDFs stable for any image pairs, for any signal-to-noise ratios ? What happens if $\sigma \rightarrow 0$?
  4. Neighboring patches are not independent
  5. Distortions are not Gaussian errors and spatially non-stationary ...

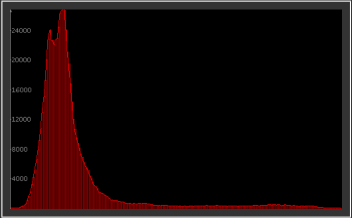
- Pros :
  1. Patch-based image representation involves intuitive algorithm parameters (patch and neighborhood sizes) and implicit regularization (patch overlapping)
  2. Collaborative neighborhood-wise decisions is attractive (e.g. statistical performance analysis)
  3. ... to be continued

Snowy traffic scene

Image 1

Image 2

difference image



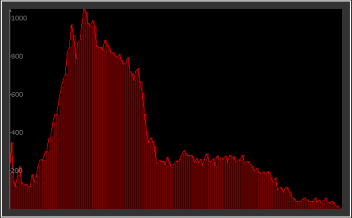
PDF of scores
(23 × 23 patches, $|\Delta(x)| = 3 \times 3$*)*

Natural outdoor scene

Image 1

Image 2

difference image



PDF of scores

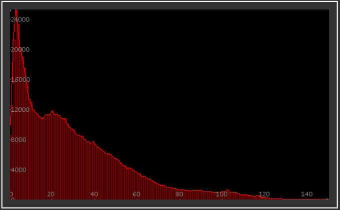
(23×23 patches, $|\Delta(x)| = 3 \times 3$)

Traffic scene

Image 1

Image 2

difference image



PDF of scores
(11×11 *patches*, $|\Delta(x)| = 5 \times 5$)

Intuitive tricks for adaptive detection ?

- **idea :** A change at a pixel in an image pair correspond to scores higher than the local highest score computed from one single image and for very small neighborhoods $\mathcal{N}(x)$ (3×3 neighborhood) ...

$$\tau(x) \triangleq \max\left(\sup_{y \in \mathcal{N}(x)} \frac{\|\underline{u}(x) - \underline{u}(y)\|^2}{2\sigma^2}, \tau_{min}\right)$$

$$\tau_{min} \triangleq \frac{1}{|\Omega|} \sum_{x \in \Omega} \inf_{y \in \mathcal{N}(x)} \frac{\|\underline{u}(x) - \underline{u}(y)\|^2}{2\sigma^2}$$

- **Maximum vote :** A change is detected at pixel x if every local score $z(x_i)$ is higher than $\tau(x)$... Setting $T = N$ implies

$$\mathbb{P}\{\text{count}(x) = N; n|H_0\} = (\mathbb{P}\{z(x_i) \geq \tau(x); n\})^N$$

Snowy traffic scene

image pair

difference image

image count(x)

change detection

23×23 *patches*, 3×3 *search windows*

thresholded difference image [Kapur 85]

Traffic scene

image pair

difference image

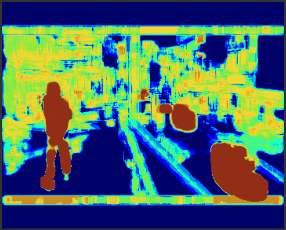


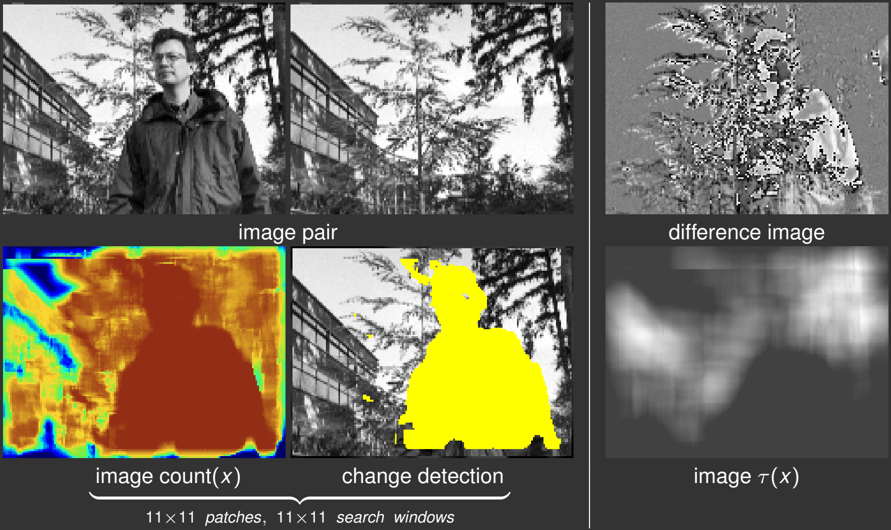
image count(x)

change detection

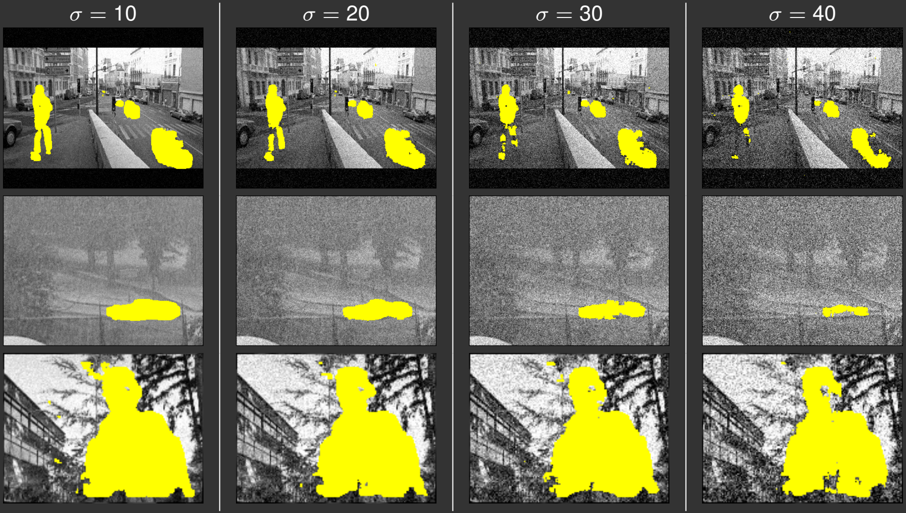
11×11 *patches*, 5×5 *search windows*

image $\tau(x)$

Outdoor scene

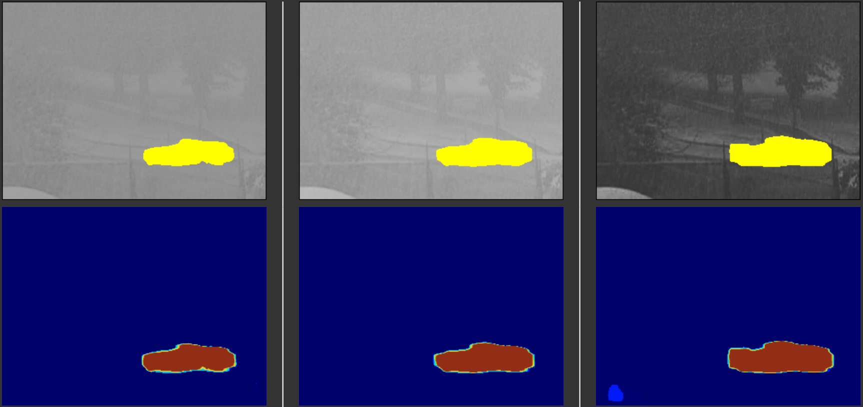


Robustness to white Gaussian noise



Robustness to contrast changes

1. Invariance to linear contrast changes applied to $w = (u, v)$
2. Robustness to moderate nonlinear contrast changes :



$w'(x) = 30\log(w(x)+1)$ | $w'(x) = 10\sqrt{(w(x)+128)}$ | $w'(x) = 10^{-5}((w(x))^3 + 50)$

Bidirectional analysis for change detection

- **Adaptive thresholds :** we consider both the two images u and v and compute the spatially varying thresholds as

$$\tau(x) \triangleq \min\left(\sup_{y\in\mathcal{N}(x)} \frac{\|\underline{u}(x)-\underline{u}(y)\|^2}{2\sigma^2}, \sup_{y\in\mathcal{N}(x)} \frac{\|\underline{v}(x)-\underline{v}(y)\|^2}{2\sigma^2}\right)$$

  to be compared to τ_{min}.

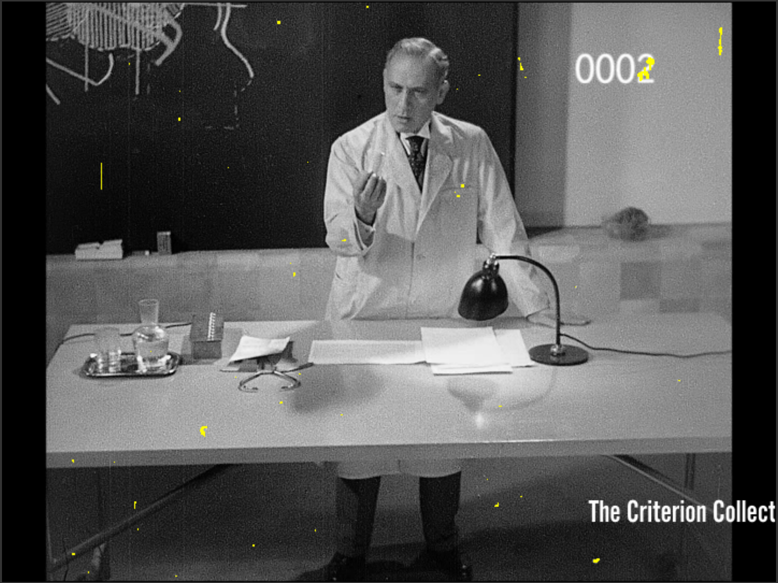
- **Local scores :**

$$z(x_i) \triangleq \min\left(\frac{\|\underline{u}(x)-\underline{v}(x_i)\|^2}{2\sigma^2}, \frac{\|\underline{v}(x)-\underline{u}(x_i)\|^2}{2\sigma^2}\right)$$

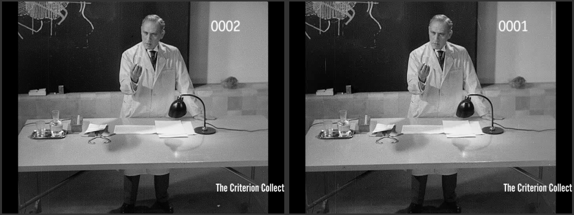
Blotches in old movies (1)

Blotches in old movies (1)

Blotches in old movies (1)



Blotches in old movies (1)



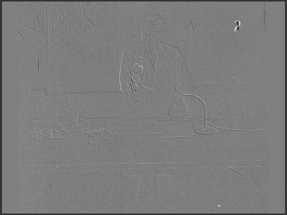


image pair

difference image

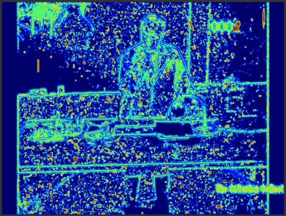


image count(x)

change detection

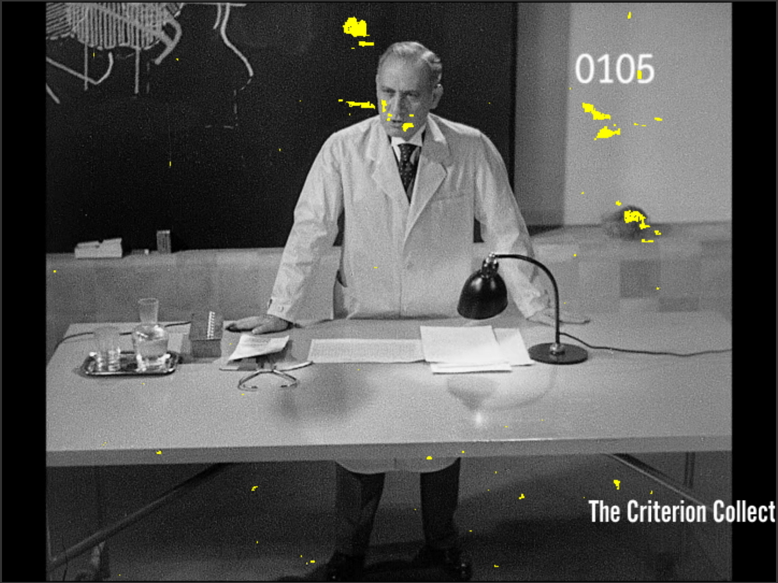
5×5 *patches*, 5×5 *search windows*

image $\tau(x)$

Blotches in old movies (2)

Blotches in old movies (2)

Blotches in old movies (2)



Blotches in old movies (2)

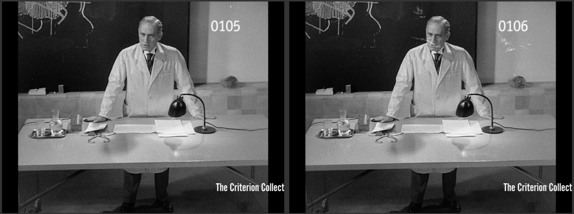
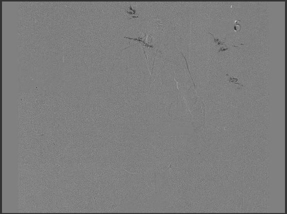


image pair



difference image

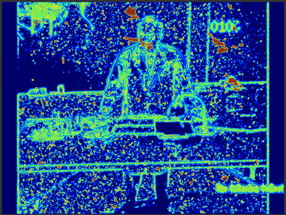
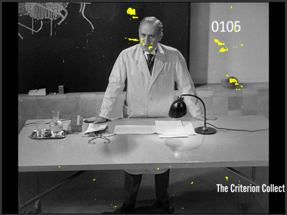


image count(x)

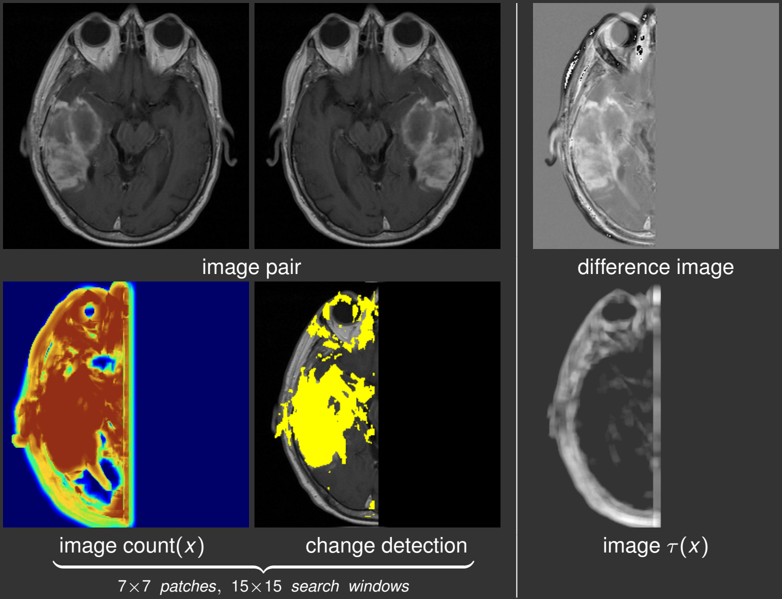


change detection

5×5 *patches*, 5×5 *search windows*

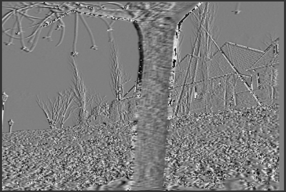
image $\tau(x)$

Asymmetries in images ?



Spatio-temporal discontinuities (1)

image pair



difference image

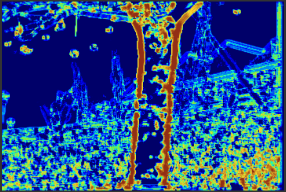


image count(x)

change detection

5×5 *patches*, 5×5 *search windows*

thresholded difference image [Kapur 85]

Spatio-temporal discontinuities (2)

3×3 patches
3×3 search windows

3×3 patches
5×5 search windows

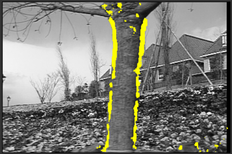
3×3 patches
7×7 search windows

5×5 patches
3×3 search windows

5×5 patches
5×5 search windows

5×5 patches
7×7 search windows

7×7 patches
3×3 search windows

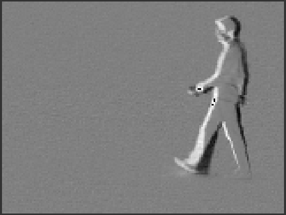


7×7 patches
5×5 search windows

7×7 patches
7×7 search windows

Space-time interest points ?

image pair



difference image

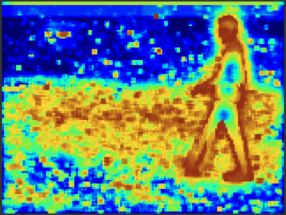


image count(x)

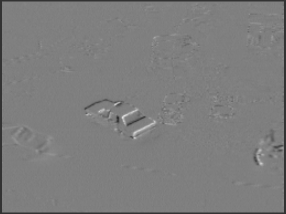
space-time interest points

3×3 *patches*, 15×15 *search windows*

thresholded difference image [Kapur 85]

Space-time interest points ?


image pair



difference image

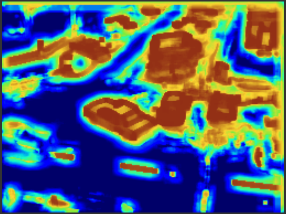


image count(x)

space-time interest points

3×3 *patches*, 15×15 *search windows*

thresholded difference image [Kapur 85]

Patch-space for multiple detection analysis

- **Consistency and patch size :** The number of false alarms can be reduced using a "scale-space" analysis.
- **Naive analysis :** A change occurs at pixel x if the number of detection for different patch sizes is "meaningful", that is

$$\begin{aligned}\mathbb{P}\left(\sum_{l=1}^{L} D_l(x) \geq k_D(x) \mid H_0\right) &= B\left(\frac{1}{L|\Omega|}\sum_{x\in\Omega}\sum_{k=1}^{L} D_l(x), L, k_D(x)\right) \\ &\leq \frac{\varepsilon}{|\Omega|}\end{aligned}$$

where $B(\cdot)$ is the tail of the Binomial distribution, L is the number of patch sizes, $k_D(x)$ is the number of positive decisions in the collection $\mathcal{D}(x) = \{D_1(x), \ldots, D_L(x)\}$ and ε is the expected number of false alarms in the "scale-space" volume.

Patch similarity and invariance

- **Euclidean distance :**

$$z_E(x) = \|\underline{u}(x) - \underline{v}(y)\|^2$$

- **Illumination invariance :**
  1. **Local contrast changes**

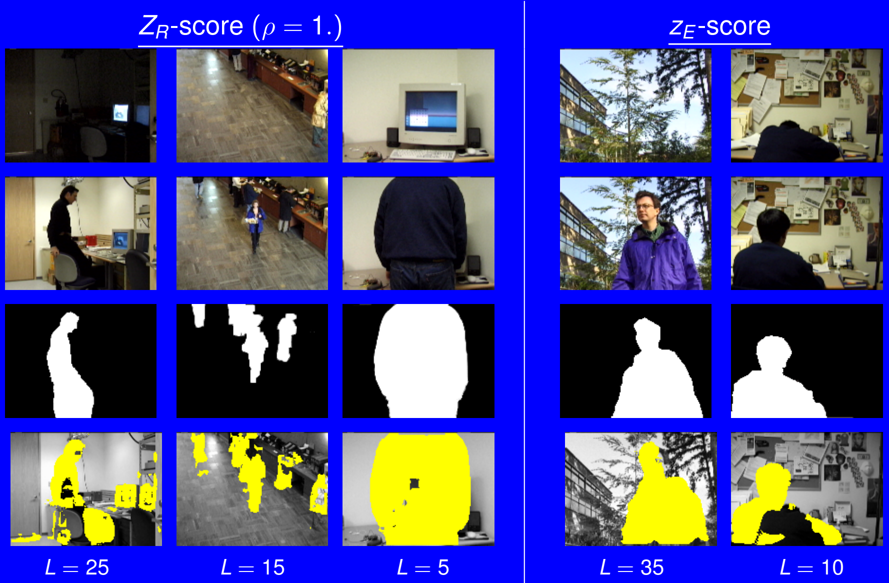
$$z_C(x) = \|\underline{u}(x) - \underline{v}(y) - (u_\rho(x) - v_\rho(y))1_n\|^2$$

  2. **Specularity and shadow effects**

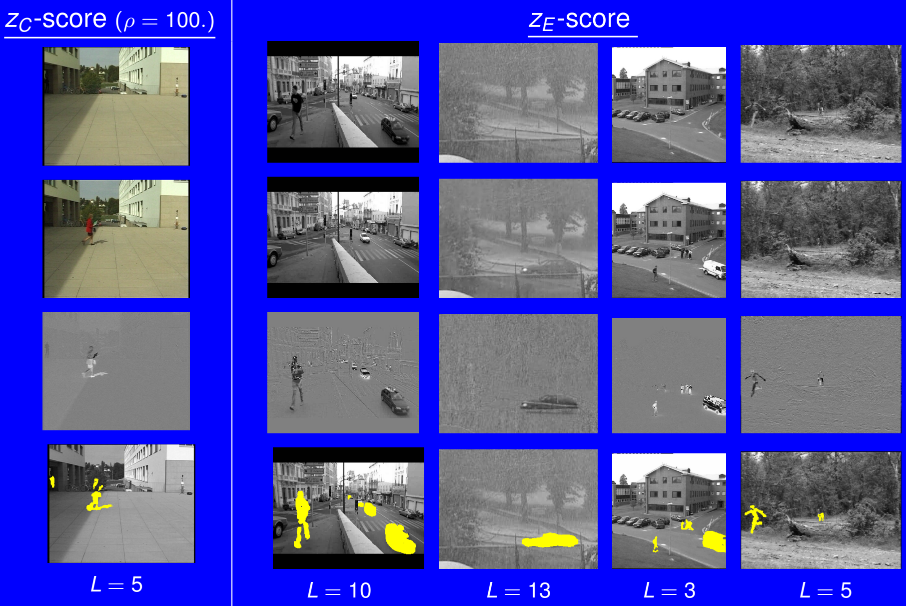
$$z_R(x) = \|\underline{u}(x) - \frac{u_\rho(x)}{v_\rho(y)}\underline{v}(y)\|^2$$

where 1_n is a 1-vector with n elements and $u_\rho = G_\rho \star u$ and $v_\rho = G_\rho \star v$ are the images convolved with a Gaussian kernel G_ρ with standard deviation ρ.

- **Global motion compensation for affine motion invariance**
(Odobez & Bouthemy, JVCIR95)



First rows : image pairs (160 $\times$ 120 pixels) ; Third row : ground truth images ;
Fourth row : our detection results with the z_R-score and the z_E-score.



First rows : image pairs ; Third row : difference images ; Fourth row : our detection results.

Conclusion & Perspectives

1. **Summary :**
   - Patch-based image representation
   - Detection of changes, occlusions and space-time corners
   - Intuitive algorithm parameters and regularization
   - Collaborative neighborhood-wise decisions
   - Performance analysis (false alarm probabilities)

2. **Perspectives :**
   - More experimental results
   - Conditional Random Fields modeling and global optimization (Graph Cuts) ... to get similar results ?