Robust Visual Correspondence: Theory and Applications

Federico Tombari, Luigi Di Stefano and Stefano Mattoccia
DEIS/ARCES – University of Bologna
Italy

1. Introduction

Visual correspondence represents one of the most important tasks in computer vision. Given two sets of pixels (i.e. two images), it aims at finding corresponding pixel pairs belonging to the two sets (*homologous* pixels). As a matter of fact, visual correspondence is commonly employed in fields such as stereo correspondence, change detection, image registration, motion estimation, pattern matching, image vector quantization.

The visual correspondence task can be extremely challenging in presence of disturbance factors which typically affect images. A common source of disturbances can be related to photometric distortions between the images under comparison. These can be ascribed to the camera sensors employed in the image acquisition process (due to dynamic variations of camera parameters such as auto-exposure and auto-gain, or to the use of different cameras), or can be induced by external factors such as changes of the amount of light emitted by the sources or viewing of non-lambertian surfaces at different angles.

All of these factors tend to produce brightness changes in corresponding pixels of the two images that can not be neglected in real applications implying visual correspondence between images acquired from different spatial points (e.g. stereo vision) and/or different time instants (e.g. pattern matching, change detection). In addition to photometric distortions, differences between corresponding pixels can also be due to the noise introduced by camera sensors. Finally, the acquisition of images from different spatial points or different time instants can also induce occlusions. Evaluation assessments have also been proposed which compared visual correspondence approaches for tasks such as stereo correspondence (Chambon & Crouzil, 2003), image registration (Zitova & Flusser, 2003) and image motion (Giachetti, 2000).

2. Literature review

Let *Ir, It* be respectively the reference image patch vector and the target image patch vector, that have to be matched together. Traditional matching measures can be subdivided into two categories: correlation-based or distance-based. Between the first, the most commonly adopted are the *Normalized Cross-Correlation* (NCC) and the *Zero-mean Normalized Cross-Correlation* (ZNCC):

$$\text{NCC}(I_r, I_t) = \frac{I_r \circ I_t}{\|I_r\|_2 \cdot \|I_t\|_2} \qquad (1)$$

$$ZNCC(I_r, I_t) = \frac{(I_r - \overline{I_r}) \circ (I_t - \overline{I_t})}{\left\| I_r - \overline{I_r} \right\|_2 \cdot \left\| I_t - \overline{I_t} \right\|_2} \tag{2}$$

with $\circ$ denoting the dot product, $|| \cdot ||_p$ the L_p norm, $\overline{}$ the mean value over the patch. Thanks to normalization with regards to the magnitude of the vectors and to the mean intensity value of the image patch, NCC and ZNCC are invariant, respectively, to linear and affine transformation between *Ir* and *It*. Efficient techniques for exhaustive template matching based on NCC and ZNCC matching measures have been proposed in (Mattoccia et al., 2008-1) and (Mattoccia et al., 2008-2).

On the other side, commonly used dissimilarity measures are those derived from the L_p-distance between *Ir* and *It*. Between this class, two popular measures are the *Sum of Absolute Differences* (SAD) and the *Sum of Squared Differences* (SSD):

$$SAD(I_r, I_t) = \left\| I_r - I_t \right\|_1 \tag{3}$$

$$SSD(I_r, I_t) = \left\| I_r - I_t \right\|_2^2 \tag{4}$$

These two measures showed experimentally good robustness towards noise (Aschwanden & Guggenbuhl, 1992), (Martin & Crowley, 1995). An efficient technique for exhaustive template matching based on L_p-distance has been proposed in (Tombari et al., 2008). A similar approach was also adopted in (Mattoccia et al., 2007) for motion estimation.

While all these measures are usually computed directly on the pixel intensities of the images, in (Martin & Crowley, 1995) it was shown that by computing these measures on the gradient norm of each pixel a higher robustness is attained, i.e. for what concerns insensitivity to illumination changes the SSD and the NCC applied on gradient norms (referred to here respectively as G-SSD and G-NCC) showed to perform well. In particular, if we denote with $G_r(i, j)$ the gradient of *Ir* at pixel (i, j):

$$G_r(i, j) = \left[\frac{\partial I_r(i, j)}{\partial i}, \frac{\partial I_r(i, j)}{\partial j} \right]^T = \left[G_i^r(i, j), G_j^r(i, j) \right]^T \tag{5}$$

and similarly with $G_t(i, j)$ the gradient of *It* at pixel (i, j):

$$G_t(i, j) = \left[\frac{\partial I_t(i, j)}{\partial i}, \frac{\partial I_t(i, j)}{\partial j} \right]^T = \left[G_i^t(i, j), G_j^t(i, j) \right]^T \tag{6}$$

the gradient norm, or magnitude, in both cases is defined as:

$$\left\| G_r(i, j) \right\|_2 = \sqrt{G_i^r(i, j)^2 + G_j^r(i, j)^2} \tag{7}$$

$$\left\| G_t(i, j) \right\|_2 = \sqrt{G_i^t(i, j)^2 + G_j^t(i, j)^2} \tag{8}$$

Hence the G-NCC function can be defined as:

$$G\text{-}NCC(I_r, I_t) = \frac{\sum_{(i,j) \in I_r} \left\| G_r(i, j) \right\|_2 \cdot \left\| G_t(i, j) \right\|_2}{\sqrt{\sum_{(i,j) \in I_r} \left\| G_r(i, j) \right\|_2^2} \cdot \sqrt{\sum_{(i,j) \in I_t} \left\| G_t(i, j) \right\|_2^2}} \tag{9}$$

and, analogously, the G-SSD function as:

$$\text{G-SSD}(I_r, I_t) = \sum_{(i,j) \in I_r} \left(\left\| G_r(i,j) \right\|_2 - \left\| G_t(i,j) \right\|_2 \right)^2 \tag{10}$$

In addition to these measures, many alternatives have been proposed in literature with the specific aim of deploying robust image matching. The *Gradient Correlation* (GC) measure, proposed in (Crouzil et al., 1996) and derived from a measure originally introduced in (Scharstein, 1994), is based on two terms, referred to as *distinctiveness* (D) and *confidence* (C), both computed from intensity gradients:

$$D(I_r, I_t) = \sum_{(i,j) \in I_r} \left\| G_r(i,j) - G_t(i,j) \right\|_2 \tag{11}$$

$$C(I_r, I_t) = \sum_{(i,j) \in I_r} \left(\left\| G_r(i,j) \right\|_2 + \left\| G_t(i,j) \right\|_2 \right) \tag{12}$$

The GC measure is then defined as:

$$GC(I_r, I_t) = \frac{D(I_r, I_t)}{C(I_r, I_t)} \tag{13}$$

Its minimum value is 0, corresponding to the highest similarity between *Ir* and *It*. For any other positive value, the greater the value, the higher the dissimilarity between the two vectors. In order to compute the partial derivatives, (Crouzil et al., 1996) proposes to use either the Sobel operator or the Shen-Castan ISEF filter (Shen & Castan, 1992).

The *Orientation Correlation* (OC) measure (Fitch et al., 2002) is based on the correlation of the orientation of the intensity gradient. In particular, for each gradient $G_r(i,j)$ a complex number representing the orientation of the gradient vector is defined as:

$$O_r(i,j) = \text{sgn}\left(G_i^r(i,j) + i G_j^r(i,j) \right) \tag{14}$$

with *i* denoting the imaginary unit and where:

$$\text{sgn}(x) = \begin{cases} 0 & if \ |x| = 0 \\ \dfrac{x}{|x|} & elsewhere \end{cases} \tag{15}$$

Analogously, a complex number representing the orientation of the gradient vector $G_t(i,j)$ is defined as:

$$O_t(i,j) = \text{sgn}\left(G_i^t(i,j) + i G_j^t(i,j) \right) \tag{16}$$

As proposed in (Fitch et al., 2002), the partial derivatives for the gradient computation should be calculated by approximating them with *central differences*. Hence, the OC measure between *Ir* and *It* is defined as the real part of the correlation between all gradient orientations belonging to *Ir* and *It*:

$$OC(I_r, I_t) = \text{Re}\left\{\sum_{(i,j) \in I_r} O_r(i,j) \cdot O_t^*(i,j)\right\} \qquad (17)$$

with * indicating the conjugate of the complex vector. In (Fitch et al., 2002) it is proposed to exploit the correlation theorem to compute the correlation operation in the frequency domain by means of the FFT in order to achieve computational efficiency.

Another class of measures concerns the so-called order-consistency or order-preservation hypothesis, that is the assumption that the considered distortions do not violate the ordering between the intensities of neighbouring pixels. This assumption includes a more general class of transformations compared to the linear or affine case. These measures are called *ordinal* and a typical example of this class is represented by the *Rank transform*. As for this measure, both *Ir* and *It* are transformed into two novel images where each pixel stores the number of points in the patch whose intensity is less than that of the central point of the patch:

$$R_r(i,j) = \left|\{(u,v) \in I_r \mid I_r(u,v) < I_r(i,j)\}\right| \qquad (18)$$

$$R_t(i,j) = \left|\{(u,v) \in I_t \mid I_t(u,v) < I_t(i,j)\}\right| \qquad (19)$$

where $|\cdot|$ represents the cardinality operator. Once the two transforms are computed, a matching measure is deployed to compare *Rr* and *Rt*, e.g. (Zabih & Woodfill, 1994) proposes to use the SAD.

A typical example of this class is represented by the *Rank transform* (Zabih & Woodfill, 1994), and the measure proposed in (Bhat & Nayar, 1998). Further approaches of robust visual correspondence measures specifically conceived for change detection are (Ohta, 2001), (Xie at al., 2004), (Mittal & Ramesh, 2006).

Finally, other robust approaches have been proposed in (Seitz, 1989), (Lai, 2000), (Odone et al., 2001), (Ullah et al., 2001), (Kaneko et al., 2003).

3. The MF measure

This section describes a novel approach, referred to here as *Matching Function* (MF), which is implicitly based on the ordering assumption. In particular, MF aims at quantifying how well the order is preserved between corresponding pairs of neighbouring pixels in the two images. A simple and effective approach for evaluating the order-consistency is to evaluate the difference between the intensities of pairs of neighbouring pixels. As an example, let *Ir* be a 3×3 patch. In order to evaluate the order preservation between neighbouring elements within this window, many pairs (e.g. 72) should be considered, as each of the 9 pixels has to be put in correspondence with each other. In order to simplify the problem, we propose to consider only a subset of the whole neighbouring pairs set by evaluating only horizontal and vertical neighbouring pixels. Hence, the considered pairs are reduced to 18, as shown in Fig. 1.

In particular, in order to quantify how well the ordering is preserved between the two image patches *Ir* and *It* we propose to correlate the differences between the considered corresponding pairs within the 3×3 window. If the ordering is preserved for a given pair, the result of the pointwise correlation is a positive coefficient regardless of the sign of the

intensity difference, which tends to increase the correlation score associated with the 3 × 3 window. Conversely, if the order is not preserved the correlation coefficient is negative, and the correlation score is decreased. Moreover, since horizontal and vertical differences may be thought as the discrete approximation of the horizontal and vertical derivatives of the image, the proposed measure can also be interpreted as the cross-correlation between two vectors made out of derivatives computed within the two 3×3 patches.
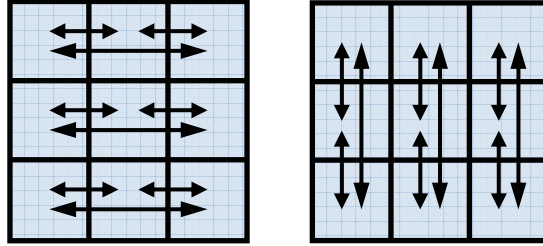

Fig. 1. considered subset of horizontal and vertical pairs of neighbouring pixels in a 3x3 patch

In the general case of two MxN patches, the considered pairs of pixels in each set include all pixels at distance 1 and 2 along horizontal and vertical directions. In order to compute this set, we define a vector of pixel differences computed at a point (i, j) on Ir:

$$\delta_{1,2}^{r}(i, j) = \begin{bmatrix} I_r(i-1, j) - I_r(i, j) \\ I_r(i, j-1) - I_r(i, j) \\ I_r(i-1, j) - I_r(i+1, j) \\ I_r(i, j-1) - I_r(i, j+1) \end{bmatrix} \tag{20}$$

and, similarly, at a point (i, j) on It:

$$\delta_{1,2}^{t}(i, j) = \begin{bmatrix} I_t(i-1, j) - I_t(i, j) \\ I_t(i, j-1) - I_t(i, j) \\ I_t(i-1, j) - I_t(i+1, j) \\ I_t(i, j-1) - I_t(i, j+1) \end{bmatrix} \tag{21}$$

Hence, the MF function consists in correlating these two vectors for each point of Ir, It, and in normalizing the correlation with the $L2$ norm of the vectors themselves:

$$MF_{1,2}(x, y) = \frac{\sum_{(i,j)\in I_r} \delta_{1,2}^{r}(i, j) \circ \delta_{1,2}^{t}(i, j)}{\sqrt{\sum_{(i,j)\in I_r} \delta_{1,2}^{r}(i, j) \circ \delta_{1,2}^{r}(i, j)} \cdot \sqrt{\sum_{(i,j)\in I_r} \delta_{1,2}^{t}(i, j) \circ \delta_{1,2}^{t}(i, j)}} \tag{22}$$

It is worth noticing that the normalization allows the measure to range between [−1,1]. It is a peculiarity of this method that, because of the correlation between differences of pixel pairs, intensity edges tend to determine higher correlation coefficients (in magnitude) with respect

to low-textured regions. Thus, this can be seen as if the measure mostly relies on the patch edges. For this reason, MF can be usefully employed also in presence of high levels of noise, as this disturbance factor can typically violate the ordering constraint on low-textured regions, but seldom along intensity edges. Similar considerations can be made in presence of partially occluded patches.

The set of pixel pairs in (21, 22) can be seen as made out of two subsets: the set of horizontal and vertical lateral derivatives (i.e. all pixels at distance 1 one to another along horizontal and vertical directions), and the set of horizontal and vertical central derivatives (i.e. all pixels at distance 2 one to another along same directions). Theoretically, the former should benefit of the higher correlation given by adjacent pixels, while the latter should be less influenced by quantization ("sampling") noise that is introduced by the camera sensor. We will refer to two additional measures of the MF class applied on each of these two subsets as, respectively, *MF*1 and *MF*2. For these last two cases, we define the vector of pixel differences at distance 1 pixel:

$$\delta_1^r(i, j) = \begin{bmatrix} I_r(i-1, j) - I_r(i, j) \\ I_r(i, j-1) - I_r(i, j) \end{bmatrix} \tag{23}$$

$$\delta_1^t(i, j) = \begin{bmatrix} I_t(i-1, j) - I_t(i, j) \\ I_t(i, j-1) - I_t(i, j) \end{bmatrix} \tag{24}$$

and the pixel differences relative to the case of distance 2 pixels:

$$\delta_2^r(i, j) = \begin{bmatrix} I_r(i-1, j) - I_r(i+1, j) \\ I_r(i, j-1) - I_r(i, j+1) \end{bmatrix} \tag{25}$$

$$\delta_2^t(i, j) = \begin{bmatrix} I_t(i-1, j) - I_t(i+1, j) \\ I_t(i, j-1) - I_t(i, j+1) \end{bmatrix} \tag{26}$$

Then, *MF*1 and *MF*2 are defined respectively as:

$$MF_1(x, y) = \frac{\displaystyle\sum_{(i,j)\in I_r} \delta_1^r(i, j) \circ \delta_1^t(i, j)}{\sqrt{\displaystyle\sum_{(i,j)\in I_r} \delta_1^r(i, j) \circ \delta_1^r(i, j)} \cdot \sqrt{\displaystyle\sum_{(i,j)\in I_r} \delta_1^t(i, j) \circ \delta_1^t(i, j)}} \tag{27}$$

and:

$$MF_2(x, y) = \frac{\displaystyle\sum_{(i,j)\in I_r} \delta_2^r(i, j) \circ \delta_2^t(i, j)}{\sqrt{\displaystyle\sum_{(i,j)\in I_r} \delta_2^r(i, j) \circ \delta_2^r(i, j)} \cdot \sqrt{\displaystyle\sum_{(i,j)\in I_r} \delta_2^t(i, j) \circ \delta_2^t(i, j)}} \tag{28}$$

A graphical representation of the 3 different pixel pair sets used by $MF_{1,2}$, MF_1 and MF_2 is shown in Fig. 2.
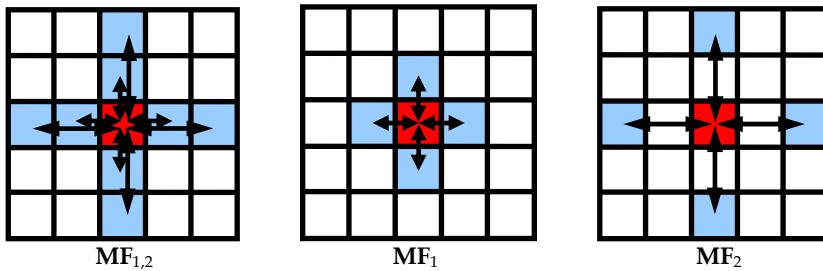
Fig. 2. The 3 considered sets of neighbouring pixel pairs.

4. Application to pattern matching

This section shows the application of the class of measures referred to as MF in a typical pattern matching scenario. Pattern matching aims at finding the most similar instances of a given pattern, *P*, within an image. In particular, in this section MF measures are compared against traditional general purpose approaches as well as against proposals specifically conceived to achieve robustness. One goal of the proposed comparison is to determine which measure is more suitable to deal with the aforementioned disturbance factors represented by photometric distortions, noise and occlusions.

More precisely, in the comparison with MF we will consider the following matching measures: GC (Crouzil et al., 1996), OC (Fitch et al., 2002), G-NCC, G-SSD (Martin and Crowley, 1995). Considered traditional measures are NCC, ZNCC and SSD. All the considered measures are tested on 3 datasets which represent a challenging framework for what regards the considered distortions. These datasets, which are publicly available at www.vision.deis.unibo.it/pm-eval.asp, are characterized by a significant presence of the disturbance factors discussed previously, and are now briefly described.

Guitar. In this dataset, 7 patterns were extracted from a picture which was taken with a good camera sensor (3 MegaPixels) and under good illumination conditions given by a lamp and some weak natural light. All these patterns have to be sought in 10 images which were taken with a cheaper and more noisy sensor (1.3 MegaPixels, mobile phone camera). Illumination changes were introduced in the images by means of variations of the rheostat of the lamp illuminating the scene ($G1-G4$), by using a torch light instead of the lamp ($G5-G6$), by using the camera flash instead of the lamp ($G7-G8$), by using the camera flash together with the lamp ($G9$), by switching off the lamp ($G10$). Furthermore, additional distortions were introduced by slightly changing the camera position at each pose and by the JPEG compression.

Mere Poulard - Illumination Changes. In dataset *Mere Poulard - Illumination Changes* (MP-IC), the picture on which the pattern was extracted was taken under good illumination conditions given by neon lights by means of a 1.3 MegaPixels mobile phone camera sensor. This pattern is then searched within 12 images which were taken either with the same camera (prefixed by *GC*) or with a cheaper, 0.3 VGA camera sensor (prefixed by *BC*). Distortions are due to slight changes in the camera point of view and by different illumination conditions such as: neon lights switched off and use of a very high exposure time ($BC-N1$, $BC-N2$, $GC-N$), neon lights switched off ($BC-NL$, $GC-NL$), presence of structured light given by a lamp light partially occluded by various obstacles ($BC-ST1$, …,

$BC-ST5$), neon lights switched off and use of the camera flash ($GC-FL$), neon lights switched off, use of the camera flash and of a very long exposure time ($GC-NFL$). Also in this case, images are JPEG compressed.

Fig. 3. *Guitar* dataset

Mere Poulard - Occlusions. In the dataset *Mere Poulard - Occlusions* (MP-Occl) the pattern is the same as in dataset MP-IC, which now has to be found in 8 images taken with a 0.3 VGA camera sensor. In this case, partial occlusion of the pattern is the most evident disturbance factor. Occlusions are generated by a person standing in front of the camera ($OP1, \ldots, OP4$), and by a book which increasingly covers part of the pattern ($OB1, \ldots, OB4$). Distortions due to illumination changes, camera pose variations, JPEG compression are also present.

Pattern

| BC-N1 | BC-N2 | GC-N |

| GC-FL | BC-ST1 | BC-ST2 |

| BC-ST3 | BC-ST4 | BC-ST5 |

| BC-NL | GC-NL | GC-NFL |

Fig. 4. *MP-IC* dataset

The number of pattern matching instances is thus 70 for the *Guitar* dataset, 12 for the *MP-IC* dataset and 8 for the *MP-Occl* dataset, for a total of 90 instances overall. The result of a pattern matching process is considered erroneous when the coordinates of the best matching subwindow found by a certain measure are further than ±5 pixel from the correct ones.

O-P1   **O-P2**   **O-P3**

O-P4   **O-B1**   **O-B2**

O-B3   **O-B4**

Fig. 5. *MP-Occl* dataset

Figures 6 and 7 report the matching errors yielded by the considered measures respectively on each of the 3 datasets and overall. As it can be seen, approaches specifically conceived to achieve robustness generally outperform classical measures, apart from the ZNCC which performs badly in presence of occlusions but shows good robustness in handling strong photometric distortions. The two measures which yield the best performance are MF and GC, with a number of total errors respectively equal to 6 and 8. In particular, MF performs better on datasets characterized by strong photometric distortions, conversely GC seems to perform better in presence of occlusions.

For what regards the 3 MF measures themselves, it seems clear that the use of differences relative to adjacent pixels suffers of the sampling noise introduced by the camera sensor, hence they appear less reliable compared to differences computed on a distance equal to 2. Moreover, as a consequence of the fact that $MF_{1,2}$ and MF_2 yield the same results on all datasets, MF_2 seems the more appropriate measure of the class since it requires only 2 correlation terms instead of the 4 needed by $MF_{1,2}$. Finally, for what regards traditional approaches, it is interesting to note that the application of NCC and SSD on the gradient norms rather than on the pixel intensities allows for a significantly higher robustness throughout all the considered datasets.
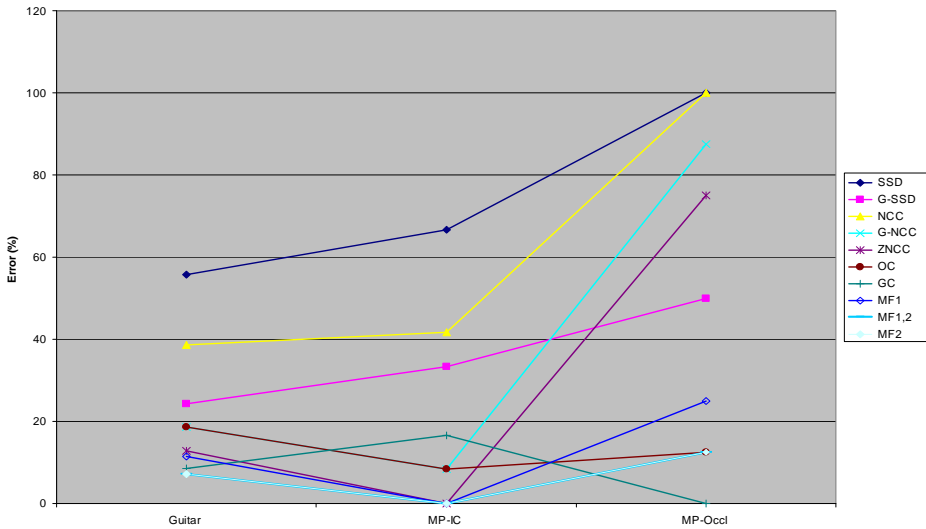
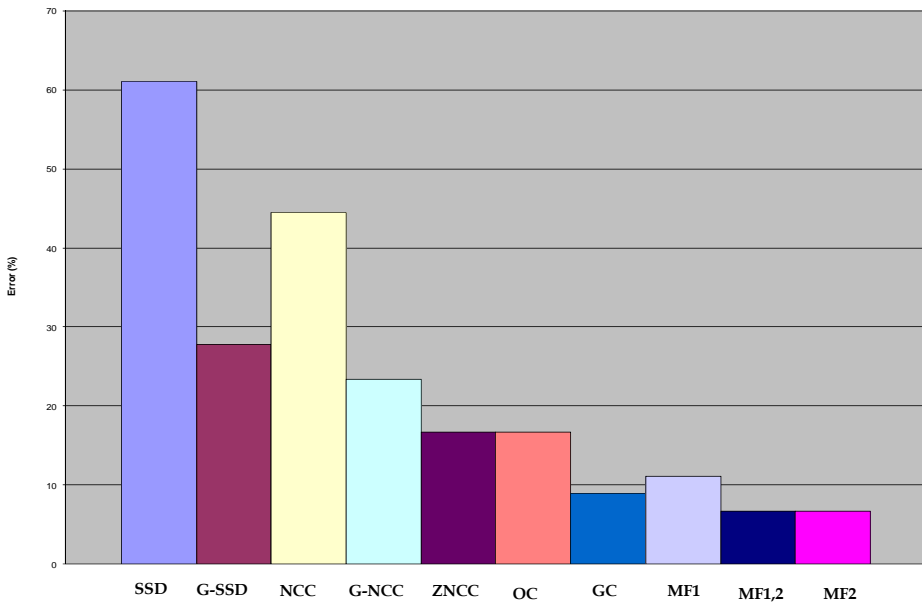Fig. 6. results of the comparison on the 3 datasets.

Fig. 7. overall results of the pattern matching evaluation.

5. Application to change detection

In this section we present the application of the proposed MF measures to the change detection task. Change detection aims at detecting structural changes occurring in time in a

scene by analyzing a sequence of frames. This is a key task in most advanced video-surveillance applications, for the mask highlighting changed pixels (*change mask*) typically represents the input data to higher level vision algorithms. This is the case of traditional single view as well as more recent and advanced multiple-views systems. The most common change detection approach is referred to as *background subtraction*: given the current frame, *F*, and a model of the background of the scene, *B*, the change mask is obtained by comparing *F* and *B*. This approach assumes that the background model is available or can be obtained by processing a short sequence of frames at initialization time. A wide variety of change detection algorithms has been proposed in literature, so as to address issues such as illumination changes, camouflage and vacillating background. A recent survey providing good coverage of this research area is given by (Radke et al., 2005).

Sudden illumination changes occurring in the scene represent a major issue for most practical change detection applications. Properly dealing with such a problem is a challenging task for change detection algorithms since the resulting photometric variations can be easily misinterpreted as structural changes, leading to many false positives in the change mask. As depicted in Fig. 8, the proposed change detection algorithm consists of three processing stages. In the first stage, the MF measure is used to extract a subset of pixels in the current frame that can be marked as background with a high confidence level. Once such a subset, referred to as F_B, is obtained, it can be usefully employed to remove the photometric distortion between *F* and *B*. To this purpose, in the second stage the algorithm computes the transformation that aligns tonally the current frame, *F*, to the background image, *B*, using as support subset F_B. In the third stage, the final change mask is achieved by a pixelwise subtraction between *F* and the tonally registered background image, B_R.
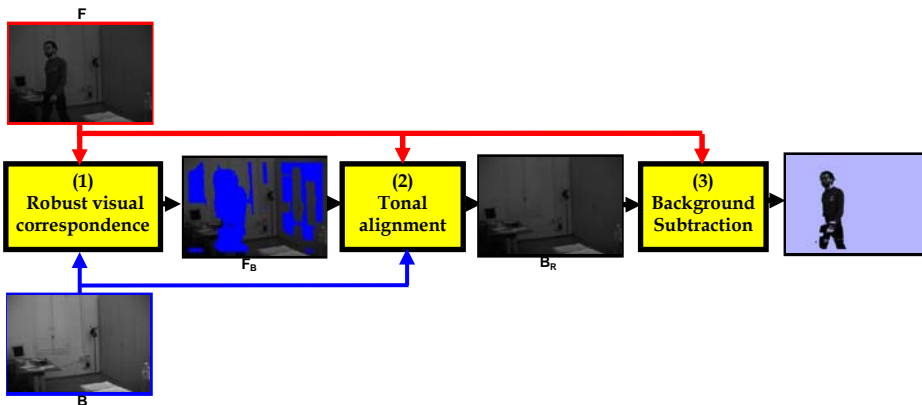

Fig. 8. Flow diagram of the proposed change detection algorithm

Robust visual correspondence In order to get F_B we match the points in the background image to the current frame. To achieve robustness with respect to outliers and noise, a block-based approach is used: that is, for each pair of correspondent points in *B* and *F*, a $M \times M$ surrounding block is considered, and the MF measure is computed between the two blocks. Points having a score higher than a given threshold are included into F_B . To explain the usefulness of the MF measure, let's discuss Fig. 9 where, for the sake of simplicity, we consider only two kind of regions, i.e. uniform and highly-textured. When dealing with a

uniform region in both *F* and *B* (case *a* in Fig. 9), photometric differences between *F* and *B* can occur due to either variations of the illumination conditions of the background scene as well as to structural changes induced by a uniform foreground object. Thus, in this case the required matching measure should yield a low score, for nothing can be said reliably on whether the point belongs to the background or not. As for cases *b,c,d*, it is easy to observe that the matching score should be low too, since there's evidence of the presence of a foreground object. Finally, when the background is highly textured and the texture pattern does not change in spite of possible photometric changes (case *e*), it is reasonable to flag the point as background with a high confidence level. Hence, in case *e* we should get a high score from the required matching measure. Based on the above considerations, we adopt the MF measure which, as previously mentioned, matches corresponding blocks of two images by implicitly checking an ordering constraint. Since photometric variations tend not to violate the ordering of intensities in a neighbourhood of pixels, MF allows handling sudden and strong illumination variations between the background scene and the current frame. As previously discussed, MF tries to match the high contrast regions (i.e. the intensity edges) of the two blocks under comparison, since only high intensity differences can provide high contributions to the correlation score. Hence, MF behaves exactly as pointed out in Fig. 9. In fact, only two highly textured and highly correlated patterns can provide a high matching score (case *e*), while the presence of at least one *untextured* region (cases *a,b,c*) or of two textured but uncorrelated patterns (case *d*) yields a low score.
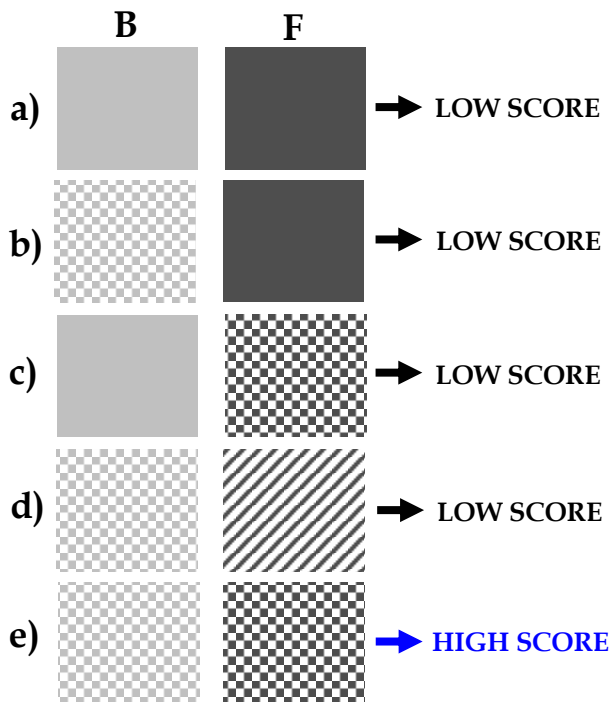

Fig. 9. Reasoning concerning the robust visual correspondence stage

Tonal alignment At this point of the algorithm, F_B represents a subset of F denoting pixels that reliably belong to the current background. Hence, B is tonally aligned to F by applying the *histogram specification* method (Gonzalez and Wood, 2002). In the evaluation of the IMF (*Intensity Mapping Function*) that aligns B to F only the set of corresponding points that belong to the mask F_B is taken into account. By applying the IMF obtained from the histogram specification method to B we get a novel background, B_R, where the photometric distortions have been removed.

Background subtraction Finally, a simple pixelwise difference between B_R and F highlights structural changes by correctly extracting foreground regions. It is worth pointing out that since background subtraction is carried out pixelwise, it is not affected by the aperture problem and allows for accurately detecting the borders as well as interior parts of foreground objects. Obviously, false negatives can still be found due to the possible camouflage between the tonally registered background and the foreground objects.

Experimental Results We now show the results dealing with a quantitative comparison between our approach and other proposals. The testing sequence was a synthetic sequence (available at: *http://muscle.prip.tuwien.ac.at/data here.php*) which comes together with groundtruth. For what regards the comparison, as representative of change detection algorithms that model false image changes according to a linear relation we consider the Normalised Cross Correlation (NCC) between pixel intensities. As for algorithms relying on checking the order preservation of intensities we consider the Rank transform (Zabih and Woodfill, 1994). We also consider as baseline for comparison the basic pixelwise background subtraction approach (BBS). For a fair comparison, we used the same block side for each algorithm (i.e. equal to 7). Then, for what regards the other parameters of each algorithm (in particular, the threshold for the final change mask), in order to determine the best parameter set of each algorithm we selected as a measure of comparison the *Precision*, i.e. the ratio between the true positives (TP) and the sum between true positives and false positives (FP), and the *Recall*, i.e. the ratio between the true positives and the sum between true positives and false negatives (FN).

In order to obtain experimental results, we started from the observation that most change detection algorithms, especially for video-surveillance applications, require to have a minimum guaranteed value of Recall. Hence, for different thresholds of minimum Recall (i.e. 70%, 80%, 90%, we selected for each algorithm the optimal parameter set maximizing the Precision value. Such results are shown in Tab. 1. It is worth pointing out that we fixed the maximum constraint value of Recall to 90%, since with higher values all algorithms would provide Precision values lower than 50%, which would result in very poor change masks (the number of false positives being higher than the number of true positives). Moreover, it is worth noting that also for these results no post processing was added to the output of the evaluated algorithms, similarly no morphology operator was used at any stage of the evaluated algorithms.

From the Table it is easy to infer that the proposed algorithm is the most robust and accurate between the evaluated ones, since it always outperforms the other approaches in terms of Precision for all different constraint values of Recall. In addition, Fig. 10 shows, for a single frame of the evaluated testing sequence, the outputs of the various algorithms at the different constraint values of Recall. In addition, in the first row of the Figure the background model as well as the current frame together with the correspondent ground truth frame are shown. These results qualitatively confirm the trend shown in Tab. 1, proving that our approach provides overall the most accurate results.

	> 70%	> 80%	> 90%
Proposed	87.3	81,7	52,2
NCC	59,6	57,2	43,0
Rank	24,5	18,8	13,1
BBS	2,2	1,9	1,7

Table 1. Best values of Precision yielded by the evaluated algorithms with different constraint values on Recall.
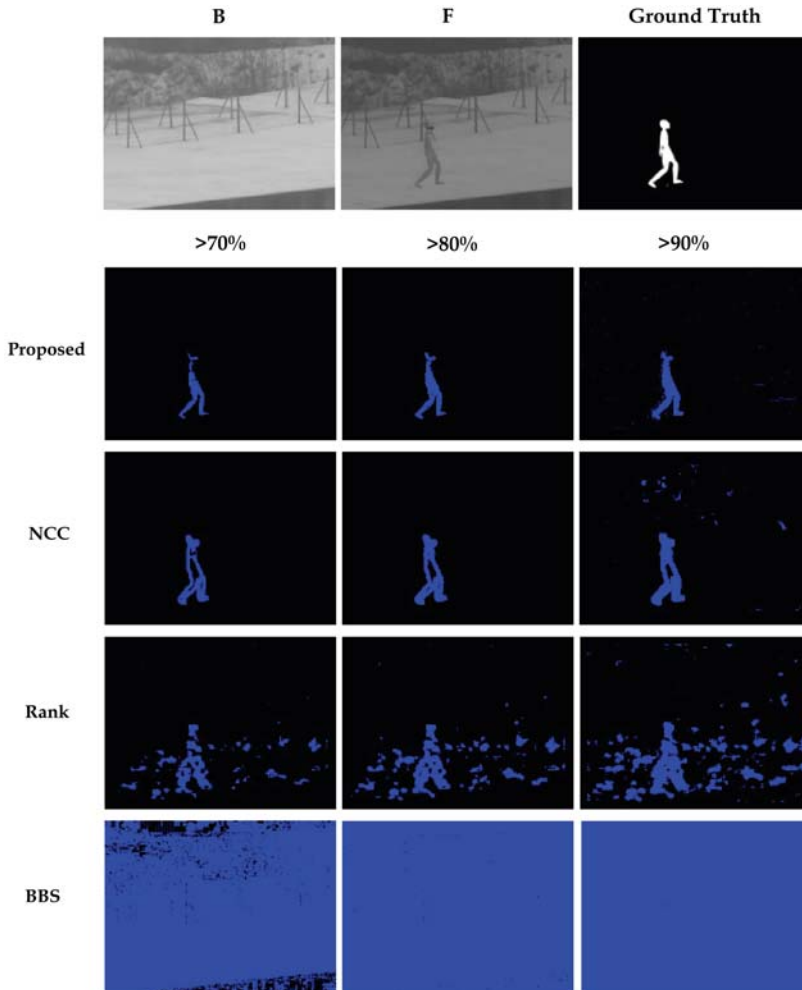

Fig. 10. Comparison of outputs yielded by the evaluated algorithms on the same sequence and with the same constraint values on Recall used for results in Tab. 1. First row, from left to right: background model *B*, current frame *F*, Ground Truth

6. Application to video-surveillance

In this section a case study is presented where access to a high security gate has to be monitored to assess for the presence/absence of people as well as to ensure that only one person occupies the gate at a given time (anti-tailgating). To accomplish this surveillance task, first of all we apply a change detection algorithm in order to discriminate between foreground (i.e. occupants) and background (i.e. the gate floor) regions of incoming frames. The change detection algorithm relies on the MF measure: at each point of the current frame and background image MF is computed on the window centred on the current point, then a threshold is used to discriminate background points from foreground points. The images are subject to heavy photometric distortions due to reflections on the gate floor, changes in indoor illumination and unpredictable light coming from outside. Fig. 11 shows the results where 3 frames acquired in different moments and with different subjects are compared with the same background image acquired previously (shown on the left). The 3 images on the right show the shape of the region detected as the gate floor using a fixed set of parameters (window side = 15, threshold = 0.2) and, as post-processing, a fixed sequence of simple morphological operators such as erosion and dilation. Results show that the proposed measure is able to extract a good shape of the gate floor with good robustness towards the ongoing disturbance factors.
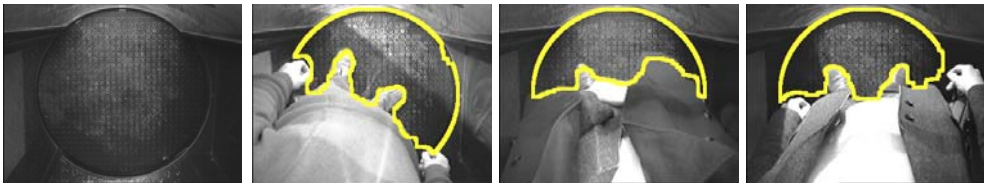

Fig. 11. The background (*left*) and 3 examples dealing with the presented video-surveillance application.

7. Conclusions

A review of the state of the art and a novel class of measures for robust visual correspondence under disturbance factors such as photometric distortions, noise and occlusions have been proposed. The proposed approach is based on the order preservation hypothesis, and aims at measuring how well the ordering constraint between neighbouring pixels is preserved. The novel measures were demonstrated to perform effectively in task such as pattern matching and change detection, as well as in a challenging surveillance scenario considered as a case study.

8. References

P. Aschwanden & W. Guggenbuhl (1992). *Experimental results from a comparative study on correlation-type registration algorithms.* In W. Forstner and S. Ruwiedel, editors, Robust computer vision, pages 268–289. Wichmann.

D. Bhat & S. Nayar (1998). *Ordinal measures for image correspondence*. IEEE Trans. Pattern Recognition and Machine Intelligence, 20(4): 415–423.

S. Chambon & A. Crouzil (2003). *Dense matching using correlation: new measures that are robust near occlusions.* Proceedings of British Machine Vision Conference, volume 1, pages 143–152.

A. Crouzil; L. Massip-Pailhes & S. Castan (1996). *A new correlation criterion based on gradient fields similarity.* In Proceedings Int. Conf. on Pattern Recognition, pages 632–636.

A. J. Fitch; A. Kadyrov; W. J. Christmas & J. Kittler (2002). *Orientation correlation.* Proceedings of British Machine Vision Conference, volume 1, pages 133–142.

S. Giachetti (2000). *Matching techniques to compute image motion.* Image and Vision Computing, 18: 247-260.

R. Gonzalez & R.Woods (2002). *Digital Image Processing.* Prentice Hall, 2nd edition.

S. Kaneko; Y. Satoh & S. Igarashi (2003). *Using selective correlation coefficient for robust image registration.*Journal of Pattern Recognition, 36(5):1165–1173.

S. Lai (2000). *Robust image matching under partial occlusion and spatially varying illumination change.* Computer Vision and Image Understanding, 78:84–98.

J. Martin & J. Crowley (1995). *Experimental comparison of correlation techniques.* Proceedings of Int. Conf. on Intelligent Autonomous Systems, volume 4, pages 86–93.

S. Mattoccia, F. Tombari & L. Di Stefano (2008-1). *Fast full-search equivalent template matching by Enhanced Bounded Correlation.* IEEE Trans. on Image Processing, 17(4): 528-538

S. Mattoccia, F. Tombari & L. Di Stefano (2008-2). *Reliable rejection of mismatching candidates for efficient ZNCC template matching.* Proceedings of IEEE Int. Conference on Image Processing (*in press*)

S. Mattoccia, F. Tombari, L. Di Stefano & M. Pignoloni (2007). *Efficient and optimal block matching for motion estimation.* 14th IAPR Int. Conference on Image Analysis and Processing, volume 1, pages 705-710

A. Mittal & V. Ramesh (2006). *An intensity-augmented ordinal measure for visual correspondence.* Proceedings of IEEE Conf. on Computer Vision and Pattern Recognition, volume 1, pages 849–856.

F. Odone; E. Trucco & A. Verri (2001). *General purpose matching of grey level arbitrary images.*Proceedings of 4th Int. Workshop on Visual Form, pages 573–582.

N. Ohta (2001). *A statistical approach to background subtraction for surveillance systems.* In Proceedings of Int. Conf. on Computer Vision, volume 2, pages 481–486.

R. Radke; S. Andra; O. Al-kofahi & B. Roysam (2005). *Image change detection algorithms: a systematic survey.* IEEE Trans. Image Processing, 14(3):294–307.

D. Scharstein (1994). *Matching images by comparing their gradient fields.* Proceedings of Int. Conf. on Pattern Recognition, volume 1, pages 572–575.

P. Seitz (1989). *Using local orientational information as image primitive for robust object recognition.* Proceedings SPIE, Visual Communication and Image Processing IV, volume 1199, pages 1630–1639.

J. Shen & S. Castan (1992). *An optimal linear operator for step edge detection.* Graphical Models and Image Processing, 54(2):112–133.

F. Tombari, S. Mattoccia & L. Di Stefano (2008). *Full search-equivalent pattern matching with Incremental Dissimilarity Approximations.* IEEE Trans. Pattern Analysis and Machine Intelligence (*in press*)

F. Ullah; S. Kaneko & S. Igarashi (2001). *Orientation code matching for robust object search.* IEICE Trans. Information and Systems, E-84-D(8):999–1006.

B. Xie; V. Ramesh & T. Boult (2004). *Sudden illumination change detection using order consistency.* Image and Vision Computing, 2(2):117–125.

R. Zabih & J. Woodfill (1994). *Non-parametric local transforms for computing visual correspondence.* In Proc. European Conf. on Computer Vision, pages 151–158.

B. Zitova & J. Flusser (2003). *Image registration methods: a survey.* Image and Vision Computing, 21(11): 977–1000.