A New Distance Measure Based on Generalized Image Normalized Cross-Correlation for Robust Video Tracking and Image Recognition

Arie Nakhmani[1,a] and **Allen Tannenbaum**[b]

[a]A. Nakhmani is with the Department of Electrical and Computer Engineering, Boston University, Boston, MA.

[b]A. Tannenbaum is with the Comprehensive Cancer Center/Department of Radiology/Department of ECE, UAB, Birmingham, AL.

Abstract

We propose two novel distance measures, normalized between 0 and 1, and based on normalized cross-correlation for image matching. These distance measures explicitly utilize the fact that for natural images there is a high correlation between spatially close pixels. Image matching is used in various computer vision tasks, and the requirements to the distance measure are application dependent. Image recognition applications require more shift and rotation robust measures. In contrast, registration and tracking applications require better localization and noise tolerance. In this paper, we explore different advantages of our distance measures, and compare them to other popular measures, including Normalized Cross-Correlation (NCC) and Image Euclidean Distance (IMED). We show which of the proposed measures is more appropriate for tracking, and which is appropriate for image recognition tasks.

Keywords

Correlation; NCC; Image distance; Template matching

1. Introduction

Many computer vision tasks require a comparison between pairs of images, where the distance measure is used to describe quantitatively how much one image is similar to another. In registration and stereo pair matching, the images are aligned to obtain the highest similarity between them. In visual tracking and robotic navigation, the current video frame is searched for the predefined target templates or features, and the location of best match is associated with the true target's location. The image can be retrieved from a large database by sending the request of matching to a given pencil sketch. Similarly, the content of an image can be extracted, analyzed, and recognized by comparing to the predefined templates. In this paper we will concentrate on the tracking and image recognition applications.

[1]Corresponding author. Address: Department of Electrical and Computer Engineering, Boston University, Boston, MA. Tel.: +1 617 358 5803; nakhmani@gmail.com..

Unfortunately, there is no a single distance measure that works well for all tasks. Different tasks demand different measure properties. For example, in image recognition we are interested to find out if the extracted object is of certain type or not. Therefore, small deformations (translations, rotations, scaling, etc.) should not change the distance a lot (if at all). In tracking, we want to localize the target, thus the distance should be much less tolerant to the translation, but robust to additive and multiplicative noise which is found in the real world scenes. But even in the tracking there may be equivocal requirements to the distance measure. If only the binary decision (target located or not) is needed, then we want the distance to be zero at the right location and the infinity otherwise, but if the distance is used to adjust the location by some iterative algorithm (from the current measurement), then we want the distance to increase gradually when moving out from the desired location. When searching for an image in the database, we want the distance to rise gradually with the enlarging deformation. In the Subsection 1.1, we will describe a few popular distance measures, their properties, and their appropriateness to the tracking and image recognition.

Image matching has a huge literature devoted to it, and so we will only be able to briefly sketch some of the most relevant methodologies for this paper. Some earlier low level image distance measures (e.g., SAD, SSD, NCC) are described in Brown [1]. The comparison of different properties of 19 similarity measures is given in Aschwanden and Guggenbuhl [2]. More recent survey of the distance and similarity measures is found in Zitova and Flusser [3]. The basic region-based image similarity measures can be divided roughly into two groups: Euclidean distance based and correlation based [2]. In many cases the correlation based measures provide superior performance [4], but they are more computationally demanding. For a popular Zero-mean Normalized Cross-Correlation (ZNCC) the methods for fast computation have been developed by Lewis [4] and recently by Yoo and Han [5]. The algorithm [5] do not use multiplications in the computing of ZNCC. To improve the image matching results, Pratt [6] uses low pass filtering of the compared images before applying the correlation, and Zhong et al. [7] proposed a nonlinear preprocessing procedure to improve the matching.

Another interesting approach by Trujillo and Izquierdo [8] is the median NCC, which uses the median function instead of the mean in ZNCC. The aforementioned correlation similarity measures are not appropriate for matching scaled or rotated targets. Scaling invariant correlation is proposed by Cahn von Seelen and Bajcsy [9]. Rotation and scaling invariant correlation is introduced in Zhao et al. [10].

Unfortunately, many of the aforementioned measures do not consider the spatial relationship between different pixels of each image, thus enlarging a deformation is not necessarily cause increasing dissimilarity, and images that look not similar to a human observer may have a high similarity measure. Moreover, gradual deformation of the image may exhibit abrupt changes in the distance or similarity measure. Our algorithms are proposed to fill the aforementioned deficiencies.

A very interesting approach to measuring distances between images was proposed by Wang et al. [11]. This distance is called IMage Euclidean Distance (IMED), and takes into account the spatial relationships between the pixels. Sun and Feng [12] proposed an algorithm for improving the computation time of IMED. This distance can be easily embedded into another high and medium level matching procedures, and it is robust to small image deformations and perturbations [11]. These properties made IMED appropriate for image recognition and visual tracking [13]. Also the generalization of IMED for tolerance to affine transform have been proposed by Liao et al. [14]. Recently, Li and Lu [15] proposed adaptive IMED, and Sun et al. [16] proposed learning IMED, which can learn the distance from the training set of images.

Inspired by the work by Wang et al. [11], we use an idea similar to correlation instead of Euclidean distance in Section 2. The general idea is that for natural images the correlation of gray-levels between the adjacent pixels is high, thus one should compute the correlation for each pixel by weighted average of correlations with the neighboring pixels. The radial weight should fade with the increasing distance from the current pixel. Moreover, the proposed similarity measure is normalized to the interval [0, 1]. We propose two different normalizations, one is based on the Normalized Cross-Correlation (NCC), and another on Zero-mean NCC (ZNCC) [2]. We show how NCC and ZNCC can be obtained from the generalized cross-correlation formulation.

In Section 3, we test the robustness of the proposed distance measures to small deformations that may occur in tracking. Regarding the tolerance to translation, rotation, scaling and additive noise, we compare our distance measures with Euclidean distance, NCC, ZNCC, and IMED. We conclude on our findings in the Section 4.

1.1. Background

In this subsection, we describe a few popular template matching techniques, and outline their properties. We suppose that both compared images are of the same size.

Let X and Y denote the intensity values of two $M \times N$ images. By lexicographic ordering, X and Y can be transformed to the $MN \times 1$ vectors x and y. Using this notation, we define four different image similarity measures.

Euclidean Distance (ED) [2]:

$$d_{ED} = \sqrt{\sum_{i=1}^{MN}(x_i - y_i)^2} \quad (1)$$

This distance is used as matching criteria: the smaller distance is an indication of the better match. In general, d_{ED} can get the values in the range $\left[0, \sqrt{MN}F\right]$, where F denotes the maximum gray-level value ($F = 255$ for regular 8 bit images). Thus, the absolute distance value is image size dependent. As in all pixel-wise distances, where the spatial connections between the pixels are not concerned, the Euclidean distance may be large for small deformations (e.g., the ED between shifted by one pixel edge images is large). In addition, ED is sensitive to noise and constant change in brightness (e.g., for small change of the brightness by 1, $Y = X + 1$, the distance is $\sqrt{MN}$).

Normalized Cross-Correlation (NCC) [2] :

$$NCC = \frac{\sum_{i=1}^{MN} x_i y_i}{\sqrt{\sum_{i=1}^{MN} x_i^2 \sum_{i=1}^{MN} y_i^2}} \quad (2)$$

The *NCC* gets the values in the interval [0, 1], where 1 indicates the best match. On the one hand, this measure is more robust than ED for noisy scenes. On the other hand, *NCC* is not invariant under the constant change in brightness. To provide this invariance, Zero-mean Cross-Correlation is used.

Zero-mean Normalized Cross-Correlation (ZNCC) [2]:

$$ZNCC = \frac{\sum_{i=1}^{MN} (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{MN}(x_i - \overline{x})^2 \sum_{i=1}^{MN}(y_i - \overline{y})^2}} \quad (3)$$

where $\overline{x}$ and $\overline{y}$ are the mean intensity values.

The distance range is the interval $[-1, 1]$ (1 for perfect match, and 0 for "no correlation"). Although *ZNCC* is invariant to constant brightness changes, but it is not defined for constant intensity images, and shows close to one correlation between approximately white and black images. This measure is known also as Pearson correlation coefficient, and is used widely in tracking applications. Note that in tracking only the positive correlation is of interest, thus max $(0, ZNCC)$ is used as the similarity measure.

All the above similarity measures and distances were pixel-wise, and ignored the spatial connection between the pixels. In contrast, the following distance measure, proposed by Wang et al. [11], explores this spatial connection, and uses it to improve the tolerance and robustness of Euclidean distance.

IMage Euclidean Distance (IMED) [11]:

$$d_{IMED} = \sqrt{\sum_{i=1}^{MN}\sum_{j=1}^{MN} g_{ij}(x_i - y_i)(x_j - y_j)} \quad (4)$$

where $G = \{g_{ij}\}_{MN \times MN}$ is symmetric positive definite matrix created by some positive definite function of the distance between the i and j pixels.

Wang et al. [11] argued that for any reasonable image distance this function should be continuous and monotonically decreasing with the increasing distance between the pixels, and proposed the following Gaussian function:

$$g_{ij} = \frac{1}{2\pi\sigma^2}\exp\left(\frac{-\text{dist}(P_i, P_j)^2}{2\sigma^2}\right) \quad (5)$$

where dist (P_i, P_j) denote the spatial distance between the pixels i and j.

With this distance measure, smaller deformation cause smaller changes in the distance. Note that in the particular case where G is the identity matrix $I_{MN \times MN}$, then $d_{IMED} = d_{ED}$, but such matrix is created not by continuous function, thus it is not IMED.

The clear advantage of IMED over the other similarity measures can be shown by comparing the distances and the correlations for two images with vertical white lines, one pixel spaced, where the second image is shifted one pixel horizontally. These images may look very similar to human. For example, for the following images:

$$\begin{pmatrix} 255 & 0 & 255 & 0 \\ 255 & 0 & 255 & 0 \\ 255 & 0 & 255 & 0 \\ 255 & 0 & 255 & 0 \end{pmatrix} \quad and \quad \begin{pmatrix} 0 & 255 & 0 & 255 \\ 0 & 255 & 0 & 255 \\ 0 & 255 & 0 & 255 \\ 0 & 255 & 0 & 255 \end{pmatrix} \quad (6)$$

The Euclidean distance gets the maximal possible value $ED = 1020$, and the Normalized Cross-Correlation is zero, and $ZNCC = -1$. In contrast, $IMED = 274$, which is more reasonable for the proposed images.

2. Image Normalized Cross-Correlation

We denote the vectorized images as in subsection 1.1 by x and y, with dimensions $MN \times 1$. Let $f:\mathbb{R}^{MN} \to \mathbb{R}^{MN}$ be continuous function, and $G = \{g_{ij}\}_{MN \times MN}$ be symmetric positive definite matrix.

The *Generalized Cross-Correlation* is defined by:

$$GCC = f(x)^T G f(y) \quad (7)$$

The NCC (2) and ZNCC (3) are particular cases of this general definition. For NCC, $G = I_{MN \times MN}$ and $f(x) = x / \sqrt{\sum_{i=1}^{MN} x_i^2}$. For ZNCC, $G = I_{MN \times MN}$ and $f(x) = (x - \overline{x}) / \sqrt{\sum_{i=1}^{MN} (x_i - \overline{x})^2}$.

We propose to investigate two more important particular cases, where G is not the identity matrix, but it is chosen similarly to the IMED matrix (5). The idea is to use locally weighted normalized cross-correlation for each pixel, where the weight is decreases with the increasing distance from the current pixel. In other words, we use all pair-wise pixel correlations in the computation, and incorporate the spatial relations in the computation. Following the discussion in [11], the angle between the metric coefficients g_{ij} is smaller than $\pi/2$, and it can be adjusted using parameter σ. Thus, the proposed measures are robust to small perturbations, and their tolerance to deformation can be adjusted.

The new similarity measures are defined as follows.

IMage Normalized Cross-Correlation (IMNCC):

$$IMNCC = \frac{\sum_{i=1}^{MN} \sum_{j=1}^{MN} g_{ij} x_i y_j}{\sqrt{\sum_{i=1}^{MN} \sum_{j=1}^{MN} g_{ij} x_i x_j \sum_{i=1}^{MN} \sum_{j=1}^{MN} g_{ij} y_i y_j}} \quad (8)$$

where g_{ij} is defined as in IMED (5).

IMage Zero mean Normalized Cross-Correlation (IMZNCC):

$$IMZNCC = \frac{\sum_{i=1}^{MN} \sum_{j=1}^{MN} g_{ij} (x_i - \overline{x})(y_j - \overline{y})}{\sqrt{s_1 s_2}} \quad (9)$$

where $\overline{x}$ and $\overline{y}$ are the mean intensity values, $s_1 = \sum_{i=1}^{MN} \sum_{j=1}^{MN} g_{ij} (x_i - \overline{x})(x_j - \overline{x})$ and $s_2 = \sum_{i=1}^{MN} \sum_{j=1}^{MN} g_{ij} (y_i - \overline{y})(y_j - \overline{y})$.

It is interesting to note that for $g_{ij} \equiv 1$ (equal weight for all pairs of pixels) there is no mutual information computed between the images, and the absolute value of both similarity

measures is identically equal to 1. This can be easily proven by separating i and j dependent variables in (8) and (9).

Lemma 1. $|IMNCC| \quad 1$ and $|IMZNCC| \quad 1.$

Proof. For IMNCC, we have to prove that

$$\left| \sum_{i=1}^{MN} \sum_{j=1}^{MN} g_{ij} x_i y_j \right|^2 \leq \sum_{i=1}^{MN} \sum_{i=1}^{MN} g_{ij} x_i x_j \sum_{i=1}^{MN} \sum_{i=1}^{MN} g_{ij} y_i y_j$$

In matrix form it can be written as [17]: $|x^T G y|^2 \quad (x^T G x)(y^T G y)$. Note that for symmetric positive definite matrix G, $x^T G y$ is a general inner product $\langle x, y \rangle$. We apply Cauchy-Schwartz inequality, which states $|\langle x, y \rangle|^2 \quad \langle x, x \rangle \langle y, y \rangle$, to this inner product, and conclude the proof for IMNCC. For IMZNCC the same proof steps can be applied to the transformed variables $x \leftarrow x - \bar{x}$ and $y \leftarrow y - \bar{y}$.

From the Lemma 1 it is clear that the distance measure associated with the IMNCC and IMZNCC, and normalized between 0 and 1, can be defined by:

$$d_{IMNCC} = 1 - \max(0, IMNCC) \quad (10)$$

and

$$d_{IMZNCC} = 1 - \max(0, IMZNCC) \quad (11)$$

2.1. Translation Analysis

Let the image y be a pure translation t of the image x. Since the metric coefficients g_{ij} are translation invariant (g_{ij} is not a function of absolute pixel locations), $\sum_i \sum_j g_{ij} x_i x_j = \sum_i \sum_j g_{ij} y_i y_i = const$, which does not depend on the translation t. In the matrix form, $IMNCC = (X^T G X)^{-1} X^T G Y$, where $(X^T G X)^{-1} X^T G$ is a constant vector of weights that does not depend on t. Thus, IMNCC of translated images can be interpreted as a weighted average of the translated image y. For simplicity of computation, let us assume that the translation and images are continuous, and $M \times N$ rectangle image is translating vertically from top to bottom. Then IMNCC is proportional to $X^T G Y$, and thus [17]:

$$IMNCC \propto \int_t^{M+t} \int_0^N \int_0^M \int_0^N e^{-((i-x)^2 + (j-y)^2)/(2\sigma^2)} djdidydx$$

$$= \sqrt{2\pi} e^{-\frac{N^2}{2\sigma^2}} \sigma^2 \left[\sigma - e^{\frac{N^2}{2\sigma^2}} \left(\sigma - N \sqrt{\frac{\pi}{2}} \mathrm{Erf}\left(\frac{N}{\sqrt{2}\sigma}\right) \right) \right] \cdot \left[\sqrt{\frac{2}{\pi}} \sigma \left(e^{-\frac{(M-t)^2}{2\sigma^2}} - 2e^{-\frac{t^2}{2\sigma^2}} + e^{-\frac{(M+t)^2}{2\sigma^2}} \right) + (M-t) \mathrm{Erf}\left(\frac{M-t}{\sqrt{2}\sigma}\right) - 2t\mathrm{Erf}\left(\frac{t}{\sqrt{2}\sigma}\right) \right.$$

For small translation $t < \sigma$, this expression is close to its maximum (good tolerance for small deformations). For $\sigma < t << M$, $IMNCC \propto M - \sqrt{\frac{2}{\pi}}\sigma e - \frac{t^2}{2\sigma^2} - t\mathrm{Erf}\left(\frac{t}{\sqrt{2}\sigma}\right)$, which is almost linear function of t (dashed line in Figure 1).

To demonstrate visually the behavior of IMNCC distance with regard to other distances, we have chosen 10 random images of the same size, and shifted them vertically from $t = 0$ to $t = 15$. Then we normalized the distances from 0 to 1. The results can be seen in Figure 1.

Similar analysis can be done for IMZNCC, where the mean is reduced from images x and y.

2.2. Noise Tolerance Analysis

Suppose that y is a noisy version of x, i.e., $y = x + \eta$, where $\eta \sim N(0, \Sigma)$ is normally distributed noise. In this case,

$$IMNCC = \frac{\sum_i \sum_j g_{ij} x_i \left(x_j + \eta_j\right)}{\sqrt{\sum_i \sum_j g_{ij} x_i x_j \sum_i \sum_j g_{ij} \left(x_i + \eta_i\right)\left(x_j + \eta_j\right)}} = \frac{\sum_i \sum_j g_{ij} x_i x_j + \sum_i \sum_j g_{ij} x_i \eta_j}{\sqrt{\sum_i \sum_j g_{ij} x_i x_j \sum_i \sum_j \left(g_{ij} x_i x_j + g_{ij} x_i \eta_j + g_{ij} \eta_i x_j + g_{ij} \eta_i \eta_j\right)}}$$

Since the noise is independent and identically distributed, and the expected correlation between the noise and image should be zero,

$$IMNCC \approx \frac{\sum_i \sum_j g_{ij} x_i x_j}{\sqrt{\sum_i \sum_j g_{ij} x_i x_j \left(\sum_i \sum_j g_{ij} x_i x_j + \sum_i g_{ii} \eta_i^2\right)}}$$

For zero noise variance, the expression is identically equal to 1 (best correlation). The term $\sum_i g_{ii} \eta_i^2$ is proportional to noise variance, and so the IMNCC will decrease with increasing noise variance, but it will decrease slower than NCC, because average correlation of random noise with image region should be lower than the correlation with a single pixel.

In the following section, we will describe representative results for IMNCC and IMZNCC measures, and we will compare their tolerance to translation, rotation, scaling, and additive noise with another measures that were discussed in the subsection 1.1.

3. Experiments and Discussion

In all the experiments the function G was chosen to be as in (5), and σ was chosen to be 1.

3.1. Tracking

In this subsection, we have chosen a noisy and low contrast video sequence of 10 frames, and manually selected the target in these frames (see Figure 2 top row). Also, we have selected randomly 10 different images of the same size from the video, one for each frame (see Figure 2 bottom row). Then we have computed the pairwise Euclidean Distance (ED), IMED, IMNCC and IMZNCC distance measures between all possible pairs of these 20 images.

The graphical result is shown in Figure 3. The distances are color coded: black is for zero distance, and white is for maximal distance. The color bar under the distances map shows the range of the distances. The left 10×10 part of the distances matrix shows the distances between the selected true targets in different frames, thus we want it to be as close to black as possible. Naturally this matrix is symmetric. The right 10×10 part shows the distances between the true targets and the random images (not targets), thus we expect this part to be as white as possible.

3.2. Tolerance to translation

In this subsection we have chosen to test the distances on a "camera man" image [18] (see Figure 4). We have taken the 15×15 sub-image (inside the rectangle) and shifted it in different directions. For each shift we have measured and compared the shifted image with the original image below it (cropped to the same size) by the distance measures described in subsection 1.1.

In order to obtain the distances of the same order, the ED and IMED were divided by $\sqrt{MNF}$ (see subsection 1.1), and NCC and ZNCC were associated with distances similarly to (10) or (11). The representative result for the vertical translation is shown in Figure 5a.

3.3. Tolerance to scale

We have resized the 15×15 sub-image from 5×5 to 25×25 sizes and computed the distances between the scaled and the centered (and cropped) original image. The results are shown in Figure 5b.

3.4. Tolerance to rotation

We have rotated the image between –45° and 45° and compared the cropped image with the 15×15 sub-image. The results are shown in Figure 5c.

3.5. Tolerance to noise

We have computed the distances between the 15×15 sub-image and the noised version of this sub-image. We have used additive Gaussian (Normally distributed) noise with the standard deviations from 0 to 90. The results are shown in the Figure 5d.

3.6. Discussion

We can see a gradual improvement in the selectivity in Figure 3 from (a) to (d), therefore IMZNCC shows the best performance in tracking. We have repeated this test with several video sequences with the similar results. The proposed distances can be easily incorporated into the high-level tracking algorithms, and they have a clear advantage of the normalized [0, 1] range, which is easy to handle. The disadvantage is in the higher computational load.

From the Figure 5 we see that IMNCC is the most tolerant to small deformations, thus it is appropriate for image recognition tasks. On the other hand, IMZNCC is less tolerant, and similarly to ZNCC is more appropriate for the localization tasks. Both IMNCC and IMZNCC has better than ZNCC robustness to noise, comparable with IMED robustness. The response to translation of IMNCC and IMZNCC is almost linear, thus intuitively reasonable gradual change in the distances may be expected when the target is slightly off-center. Regarding the rotations, the results are inconclusive. Similar results were obtained with different distance measures. This should be expected, because when the image is rotated, the pixels that are closer to the center of rotation moved less than the farther pixels, thus the spatial relations between the pixels are different for different image regions. Unlike the IMNCC and IMZNCC where we assumed "equal rights" for all image pixels, to enlarge the tolerance to rotation, one need to emphasize far from the center pixels by choosing high σ weights g_{ij}.

In addition, we have tested the same algorithms of ZNCC and IMZNCC with median used instead of the mean, as proposed by Trujillo and Izquierdo [8]. This change seems to improve the tracking results.

4. Conclusion

We have proposed a generalization to the popular Normalized Cross-Correlation similarity measures, which consider the spatial relations between the pixels in images. The IMNCC is tolerant to noise and scaling, and is almost linearly changing with translation. The conclusion is that this measure may be appropriate for image recognition applications, where small object deformations should not influence the distance. The IMZNCC has shown high selectivity in visual tracking task, where falsely detected background image patches should be separated from the true targets. Both measures have a predefined range of values

(between 0 and 1), and can be incorporated in high-level algorithms for tracking or image recognition.

Although IMNCC and IMZNCC has shown in our tests a superior performance in the matching used for tracking applications and good tolerance to small deformations, they are computationally difficult, and the algorithm for their fast computation should be developed. If more tolerance to rotation is needed, then more complex spatial relations between the pixels can be taken into account, by adjusting the weights g_{ij}. Also, the influence of median (instead the mean) on IMZNCC should be investigated more in depth.

Acknowledgments

References

1. Brown L. A survey of image registration techniques. ACM computing surveys (CSUR). 1992; 24(4):376.

2. Aschwanden P, Guggenbuhl W, Robust computer vision. Wichmann. Ch. Experimental results from a comparative study on correlation-type registration algorithms. 1992:268–289.

3. Zitova B, Flusser J. Image registration methods: a survey. Image and vision computing. 2003; 21(11):977–1000.

4. Lewis J. Fast normalized cross-correlation. Vision Interface. 1995; 10:120–123.

5. Yoo J-C, Han T. Fast Normalized Cross-Correlation. Circuits, Systems, and Signal Processing. 2009; 28(6):819–843.

6. Pratt W. Correlation techniques of image registration. IEEE transactions on Aerospace and Electronic Systems. 1974; 10:353–358.

7. Zhong S, Cao H, Zhang T. Intensity-based correlation for heterogeneous images scene matching. Proceedings of SPIE. 2007; 6786:67863B.

8. Trujillo M, Izquierdo E. A robust correlation measure for correspondence estimation. Proceedings of 2nd International Symposium on 3D Data Processing, Visualization and Transmission. 2004:155–162.

9. Cahn von Seelen, U.; Bajcsy, R. Reconnaisance, Surveillance, and Target Acquisition for the Unmanned Ground Vehicle. Vol. 1. morgan Kaufmann Publishers; 1997. Adaptive correlation tracking of targets with changing scale; p. 313-322.

10. Zhao F, Huang Q, Gao W. Image matching by normalized cross-correlation. IEEE International Conference on Acoustics, Speech and Signal Processing. 2006; 2

11. Wang L, Zhang Y, Feng J. On the euclidean distance of images. IEEE transactions on pattern analysis and machine intelligence. 2005; 27(8):1334–1339. [PubMed: 16119271]

12. Sun B, Feng J. A Fast Algorithm for Image Euclidean Distance. Chinese Conference on Pattern Recognition. 2008:1–5.

13. Mei X, Zhou S, Wu H. Integrated detection, tracking and recognition for IR video-based vehicle classification. IEEE International Conference on Acoustics, Speech and Signal Processing. 2006; 5

14. Liao M, Wang J-Y, Chen W-F, Tang Y. An a ne invariant eulidean distance between images. International Conference on Machine Learning and Cybernetics. 2006:4133–4137.

15. Li J, Lu B. An adaptive image Euclidean distance. Pattern Recognition. 2009; 42(3):349–357.

16. Sun B, Feng J, Wang L. Learning imed via shift-invariant transformation. IEEE Conference on Computer Vision and Pattern Recognition. 2009:1398–1405.

17. Strang, G. Computational Science and Engineering. Wellesley-Cambridge Press; 2007.

18. https://sites.google.com/site/nakhmania/image

- Two novel distance measures, normalized between 0 and 1, for image matching

- Advantage: Robustness of our distance measures to translation, scaling, and noise

- Comparison to other popular measures

- The first distance measure, IMNCC, is more appropriate for recognition tasks

- The second distance measure, IMZNCC, is more appropriate for visual tracking
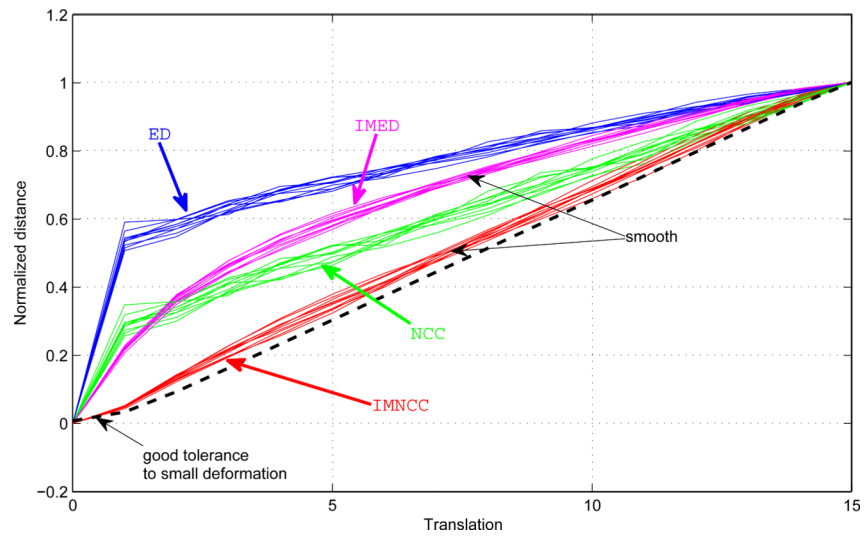
Figure 1.
Comparison of IMNCC with Euclidean Distance (ED), NCC, and IMED, with regard to translation of 10 random images. The distances are normalized between 0 and 1.
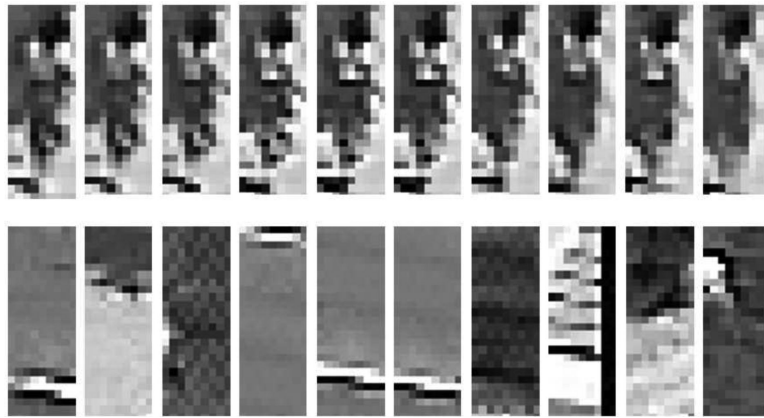
Figure 2.
(Top row) Manually selected target, walking woman, in 10 consecutive frames [18]. (Bottom row) Ten randomly selected images (same size, but different from the targets) from the same consecutive frames.
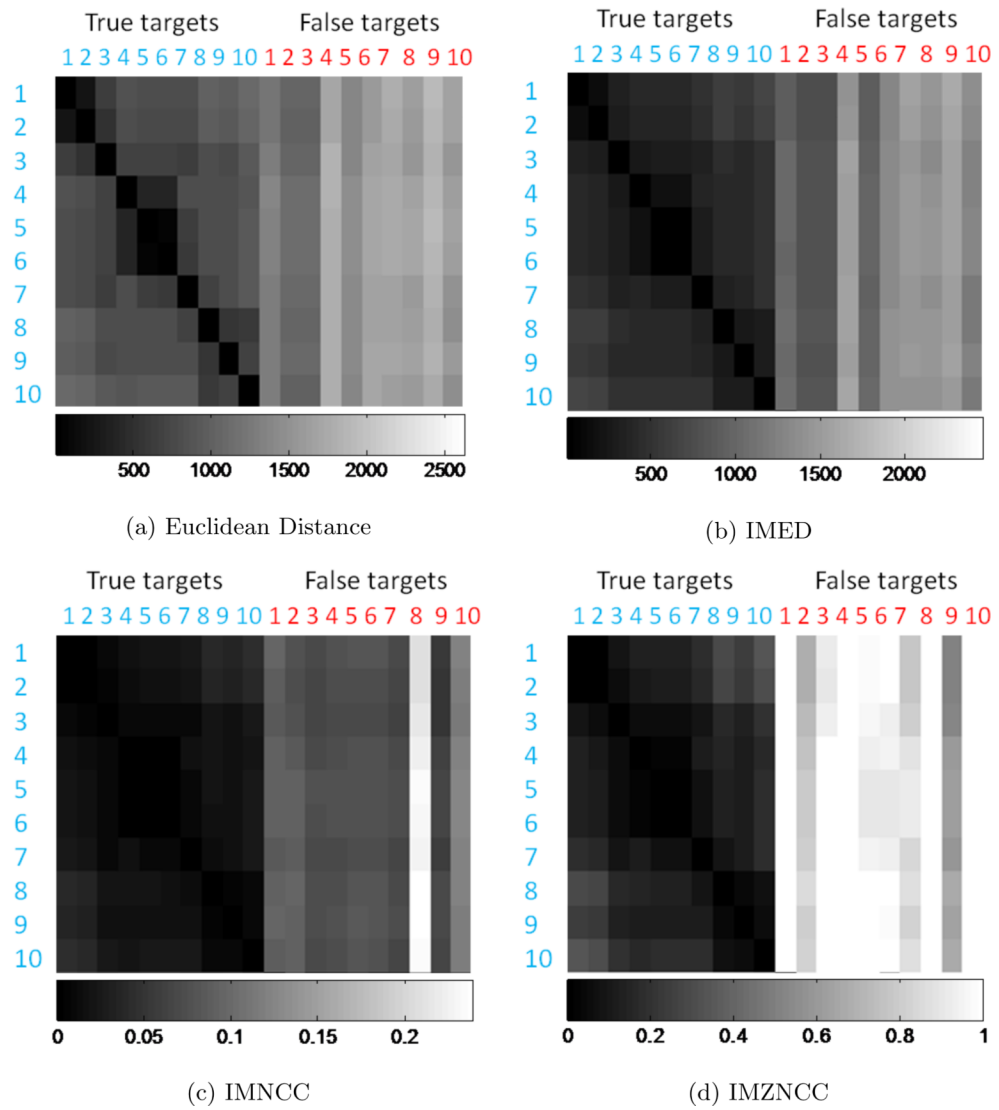
(a) Euclidean Distance

(b) IMED

(c) IMNCC

(d) IMZNCC

Figure 3.
Pairwise distances between the images in Figure 2. For the left 10×10 matrix: darker is better. For the right 10×10 matrix: brighter is better.

Figure 4.
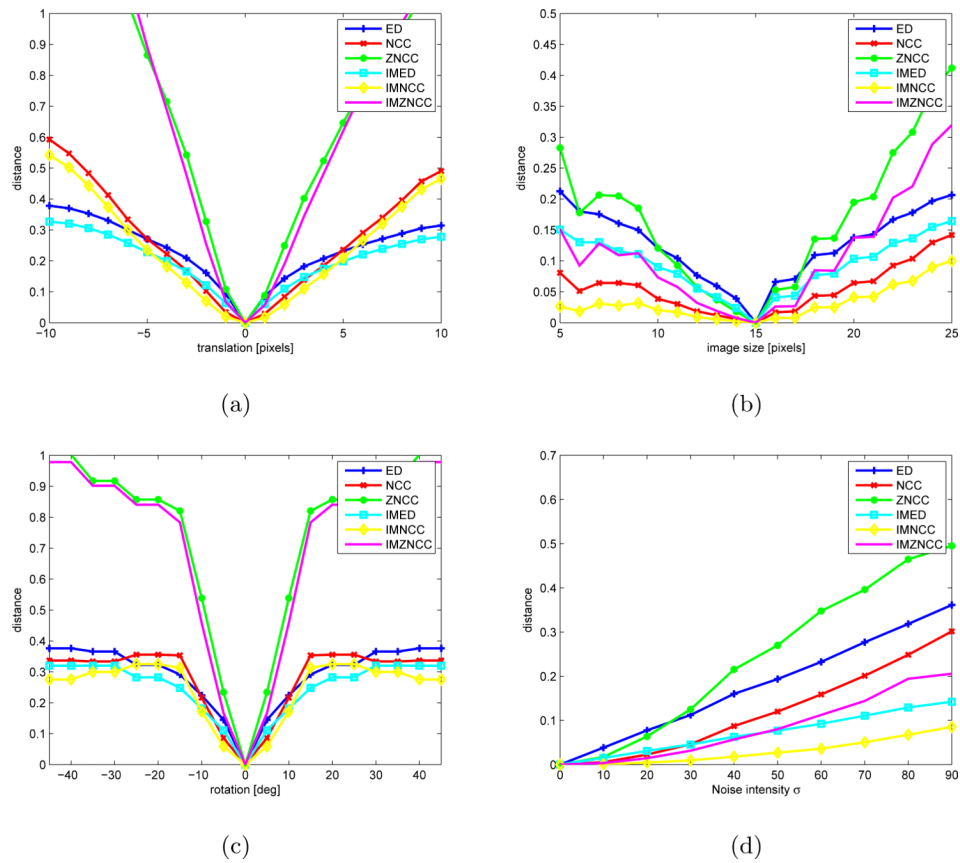Camera man image with the outlined by the rectangle tested image.

(a)

(b)

(c)

(d)

Figure 5.
(a) Distance measures comparison for vertical translation from -10 to 10 pixels. (b) Distance measures comparison for scaling from 5×5 to 25×25 image sizes (original image size is 15×15). (c) Distance measures comparison for rotation in [–45°, 45°] interval. (d) Distance measures comparison for normal noise addition up to $\sigma = 90$.