

Evaluating Robustness of Template Matching Algorithms as a Multi-objective Optimisation Problem

Jose Bernal¹, Maria Trujillo¹, and Ivan Cabezas²

¹ Universidad del Valle, Cali, Colombia

{jose.bernal,maria.trujillo}@correounivalle.edu.co

² Universidad de San Buenaventura, Cali, Colombia
imcabezas@usbcali.edu.co

Abstract. Template matching has multiple applications on different problems in computer vision. Image distortions remain as the main challenge that template matching algorithms have to overcome. Thus, measuring robustness of algorithms against distortion conditions is an important task. Moreover, a comparison among template matching algorithms is difficult to achieve due to the lack of a standard evaluation methodology. In this paper, a measurement for quantifying the robustness of template matching algorithms against a single distortion is introduced. In addition, a procedure for comparing template matching algorithms is presented, aiming to become an evaluation standard. The comparison of template matching algorithms is formulated as a Multi-objective Optimisation problem. Experimental evaluation of the proposed procedure, using the robustness coefficient, is conducted by comparing algorithms based on full-search and different similarity measurements.

Keywords: Template matching, robustness evaluation, Multiobjective Optimisation Theory.

1 Introduction

Template matching is frequently used in pattern identification, with applications in different fields, such as: medicine [1], video and image compression [13] and industry [3], among others.

Results of template matching algorithms are deviated from the expected result due to distortion in images. This deviation is also called mismatching [5]. According to the application field –in medicine, for instance–, mismatching may not be acceptable. Contributions trying to overcome image distortions consist in designing robust measures against a specific distortion, such as brightness [14], noise [11], rotation [9], scaling and translation [2]. Measures and algorithms are commonly assessed reporting: a confusion matrix [2], an interpretation of resulting values of measurements without any additional calculation [10], and an *ad-hoc* evaluation without any explanation and description of test-bed settings[11].

Proposed measures, as robust against a specific distortion, are evaluated using diverse statistics and results reported from different authors are incomparable since there is not a standard evaluation procedure.

In this paper, an evaluation procedure for comparing templates matching algorithms under distortion conditions is presented. The paper is organised as follows. A function for measuring the robustness of template matching algorithms against a distortion is introduced in Section 2. An evaluation procedure aiming to become an standard for assessing template matching algorithms is described in Section 3. The comparison of robustness exhibited by algorithms is formulated and addressed as a Multi-objective Optimisation Problem in Section 4. Experimental validation is included in Section 5. Final remarks are presented in Section 6.

2 Robustness Coefficient

Expectations of robust measures are aligned to the assertion that output tends to be similar under distortion conditions. For practical purposes, the robustness of a similarity measure against a certain image distortion is determined by the variation between the obtained best-match in the image without distortion, and the obtained best-match in the image considering a specific distortion. Under this perspective, measuring that variation is interpreted as the way for measuring robustness. A formal definition of the *robustness coefficient* is introduced:

Let A be a template matching algorithm expressed as:

$$A : \mathbb{R}^{M \times N} \times \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^2, \quad (1)$$

where $m \leq M$ and $n \leq N$.

Let $\Psi = \{\psi_0, \psi_1, \dots, \psi_d\}$ be a set of transformations from an input image into a distorted image with different severity levels varying in $[0, d]$ such that:

$$\psi_i : \mathbb{R}^{M \times N} \rightarrow \mathbb{R}^{M \times N}, \quad (2)$$

where the index $i \in [0, d]$ indicates the level of distortion. That means that ψ_0 returns an image without distortion while ψ_d returns an image with the maximum level of distortion.

Let $\iota \in \mathbb{R}^{M \times N}$ be an input image and $\tau \in \mathbb{R}^{m \times n}$ the template.

The robustness coefficient ρ of an algorithm A against the distortion Ψ , given the input image ι and the template τ , is defined by the following expression:

$$\rho(A, \Psi, \iota, \tau) = \frac{1}{|\Psi|} \sum_{i=0}^{|\Psi|} (A(\psi_0(\iota), \tau) - A(\psi_i(\iota), \tau))^2. \quad (3)$$

Thus, if the value of the robustness coefficient $\rho(\cdot)$ is close to 0, then the algorithm is robust to Ψ .

3 Evaluation Procedure

Essential elements and methods are considered in the evaluation procedure. The elements consist of a set of image distortions, input images, templates and algorithms. The methods are related to operations for performing the evaluation. The proposed procedure integrates elements and methods, as it is illustrated step by step in Figure 1.



Fig. 1. Workflow of the proposed procedure

Two fundamental considerations, for selecting image distortions and the corresponding image set, have to be taken into account: a) distortions and evaluated algorithm properties have to be aligned – for instance, it is not suitable to evaluate algorithms making pixel-by-pixel comparisons under scaling or rotating conditions –, and b) image set has to have sufficiently data quantity, gradual distortion over a same image, and annotated templates.

4 Comparison Strategy

The decision about which algorithm is the most robust against a set of distortions may not be evident as a result of the dimensionality of the problem and the limitation on optimising several objective functions simultaneously. Moreover, calculating statistic of obtained values of the robustness coefficients may not be suitable because of two main aspects. Firstly, it can be observed that the robustness coefficient range depends on the input image size. Secondly, the meaning of the robustness coefficients obtained from two distinct image distortions are different and, hence, coefficients are not comparable.

The mentioned drawbacks may be tackled as a Multi-objective Optimisation scenario. A Multi-objective Optimisation Problem (MOP) involves a large number of conflicting and inconsumable quantity of objectives. Although, a MOP is solved obtaining a Pareto Front (PF*), the analysis becomes a complex process when there are several solutions and many objectives. Thus, a method for reducing the cardinality of a PF* is introduced in [6]. A formal definition of the method is presented as follows, using the notation in Section 2.

Let $Y = \{\Psi_1, \Psi_2, \dots, \Psi_K\}$ be the set of considered distortions, $I = \{\iota_1, \iota_2, \dots, \iota_L\}$ be the set of input images, and $T = \{\tau_1, \tau_2, \dots, \tau_L\}$ be the set of corresponding templates. The comparison strategy consists in finding the algorithm $A_i \in A$ such that optimises the following equation:

$$\min_{A_i \in A} P(A_i) = \begin{bmatrix} \rho(A_i, \Psi_1, \iota_1, \tau_1) & \dots & \rho(A_i, \Psi_1, \iota_L, \tau_L) \\ \rho(A_i, \Psi_2, \iota_1, \tau_1) & \dots & \rho(A_i, \Psi_2, \iota_L, \tau_L) \\ \vdots & \ddots & \vdots \\ \rho(A_i, \Psi_K, \iota_1, \tau_1) & \dots & \rho(A_i, \Psi_K, \iota_L, \tau_L) \end{bmatrix}. \quad (4)$$

The selection of a solution from the PF* is addressed as a MOP taking into account two utility functions and the Pareto dominance relation [6]. Thus, the optimisation problem is reformulated. The problem consists in determining the vector $s = [P_{11}(\cdot), \dots, P_{KL}(\cdot)]^T$ such that minimises the following expression:

$$\min_s u(s) = [u_1(s), u_2(s)]^T, \quad (5)$$

subject to:

$$s \in PF^*, \quad (6)$$

where u_l ($l = 1, 2$) are objective functions.

Let u_1 be the sum of ranks assigned to P_{kl} in the PF*:

$$u_1(s) = \sum_{k=1}^K \sum_{l=1}^L Rank(P_{kl}(\cdot)). \quad (7)$$

Let u_2 be the sum ratios of P_{kl} in the PF*:

$$u_2(s) = \sum_{k=1}^K \sum_{l=1}^L \frac{P_{kl}(\cdot) - \min(P_{kl}(\cdot))}{\max(P_{kl}(\cdot)) - \min(P_{kl}(\cdot))}, \quad (8)$$

where $\min(P_{kl}(\cdot))$ and $\max(P_{kl}(\cdot))$, correspond to the minimum and the maximum scores of the objective at position kl , respectively. The first function measures the performance in all the involved objectives and the second refers to the best objective function values.

5 Experimental Validation

The settings of the experimental validation are presented as follows:

- Considered image distortions are: blur, brightness, low contrast, noise, and JPEG2000 compression artefact.
- Selected test-sets are *Phos* [16,15] and *CSIQ* [8] datasets. These databases contain 16 scenes of 10 original images that are gradually distorted in 6 levels. Since templates are not annotated, a manual template extraction was carried out. 16 different templates were extracted based on different sizes, textures, colors, and surfaces.
- Selected algorithms are basically formed by two main components: a search strategy and a similarity measure. Search strategy consists in an exhaustive search, called Full-Search. Measures are selected based on:
 1. *Well-known in the state-of-the-art*: MAD, MRE, MSE, SSD, SAD, and NCC [12,10]
 2. *Strength against some distortion*: ZNCC [10] and MC [14]
- Robustness coefficient $\rho(\cdot)$, in Section 2, is used for measuring algorithm robustness under a certain distortion condition and the proposed strategy, in Section 4, is used for comparing the selected algorithms once ρ values are obtained.

5.1 Assessing Robustness against a Single Distortion

Selected algorithms are evaluated against a single distortion using the robustness coefficient. The obtained results are shown in Figure 2.

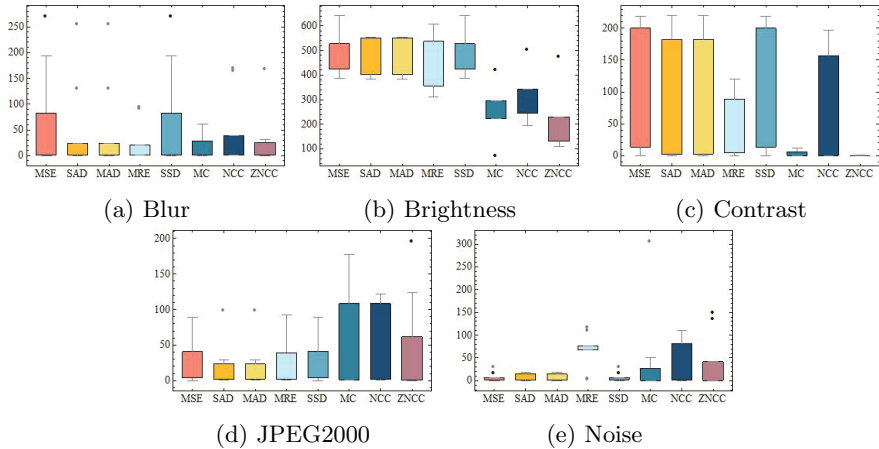


Fig. 2. Box plots of obtained results of the robustness coefficient calculated using selected image sets under different distortions

The values of the robustness coefficient calculated with the algorithm using SAD and MAD, as well as MSE and SSD are expected to be similar since one is directly proportional to the other.

The values of the robustness coefficient calculated with algorithms using different similarity measures vary regarding image distortion. The larger the IQR, the spreading of robustness coefficient values. The value of IQR in each box-plot may be an indicator of how an algorithm is affected by distortion conditions. In this sense, algorithms are highly affected under low contrast distortion since the robustness coefficient yields spread scores. In contrast, algorithms are mildly affected under noise distortion.

Measures based on differences – such as MRE, SAD, MAD, MSE, and SSD – and measures based on cross-correlation – such as MC, NCC, and ZNCC – are segmented in the evaluation of algorithms against low contrast, brightness, and noise. Therefore, brightness (illuminating conditions) may be a decisive factor in the evaluation. On the other hand, cross-correlation based measures are more affected by noise distortion than the other measures. This may be due to linear relations are affected by noisy data. A similar situation is generated when algorithms are evaluated in presence of artefacts resulting from JPEG2000 compression.

Measures based on central tendency are more robust when varying illuminating conditions, but are less robust when they are evaluated under blur, noise and artefacts generated by JPEG2000 compression.

Table 1. Algorithm ranking by distortion

Rank	Blur	Brightness	Contrast	JPEG2000	Noise
MAD	2.5	7.5	7.5	1.5	3.5
MC	5.0	2.0	2.0	8.0	5.0
MRE	1.0	3.0	3.0	3.0	7.0
MSE	7.5	5.5	5.5	4.5	1.5
NCC	6.0	4.0	4.0	7.0	8.0
SAD	2.5	7.5	7.5	1.5	3.5
SSD	7.5	5.5	5.5	4.5	1.5
ZNCC	4.0	1.0	1.0	6.0	6.0

Rankings of the robustness correlation values calculated with selected algorithms are presented in Table 1. The rank position of an algorithm varies depending on the evaluated distortion, being not a clear decision on which is the most robust algorithms.

5.2 Assessing Robustness against Multiple Distortions

Results obtained using the MOP comparison procedure, in Section 4, are compared with two strategies: Borda Count Method (BCM) [4] and Plurality Method (PM) [7]. The resulting rankings are presented in Table 2. On the one hand, the MOP strategy is more appropriate than the PM since the PM takes into account an algorithm only when obtains best robustness coefficient values while the MOP-based strategy takes into account the whole algorithm performance. On the other hand, the BCM does not take into account the relation between robustness coefficients of the same objective function such as the MOP-based strategy does with the optimisation function u_2 .

Results show that some algorithms are ranked higher than others. However, if the processing time is considered as a relevant aspect, the ranking may be different.

Table 2. Ranking of algorithms based on: MOP, BCM, and PM

Strategy	ZNCC	MC	MAD	SAD	NCC	MSE	SSD	MRE
MOP	1.0	2.0	3.5	3.5	5.0	6.5	6.5	8.0
BCM	1.0	3.0	3.0	3.0	5.0	6.5	6.5	8.0
PM	1.0	2.0	4.0	4.0	4.0	6.5	6.5	8.0

Results are easier to interpret using a MOP than analysing robustness coefficient results for each image set. In that case, the reduction of cardinality is about 88%.

6 Final Remarks

In this paper, a robustness coefficient $\rho(\cdot)$, an evaluation procedure, and a strategy for comparing template matching algorithms under distortion conditions are introduced.

The robustness coefficient is proposed in order to measure the robustness of an algorithm against a certain distortion. The measurement is based on the assumption that a robust algorithm performs in the same way even if distortion increases.

The evaluation procedure consists of ordered steps that lead to compare algorithms in the template matching context. The evaluation procedure constitutes an effort of proposing a standard for evaluating robustness of template matching algorithms. The simplicity of construction, scalability to any number of algorithms and possibility to use any method for comparing algorithms, are the main features of the suggested evaluation procedure.

A MOP-based strategy is used for comparing robustness coefficients calculated with several algorithms under different distortion conditions due to the difficulty in analysing results. The strategy is based on the analysis and reduction of the cardinality of the Pareto Front. Results show that the MOP-based strategy is easier to interpret than the initial scenario. The proposed strategy was compared with the Borda Count Method and the Plurality Method. The MOP strategy shows efficacy in handling two objectives, ranked performance of all algorithms and relation between robustness coefficients of the same objective function.

The evaluation procedure was used for comparing algorithms using differences-based measures and cross-correlation-based measures. Obtained results indicate that cross-correlation-based measures exhibit more resistance to distortions than difference-based measures. However, it is stated that rankings may vary if response time is limited.

References

1. Anumas, S., Cho, Y., Han, G., Kim, S.: Diagnosis of conductive hearing loss using template matching method. In: Biomedical Engineering International inproceedings (BMEiCON 2012), pp. 1–4 (2012)
2. de Araújo, S.A., Kim, H.Y.: Ciratefi: An rst-invariant template matching with extension to color images. *Integr. Comput.-Aided Eng.* 18(1), 75–90 (2011)
3. Arslan, O., Demirci, B., Altun, H., Tunaboyle, N.: A novel rotation-invariant template matching based on hog and amdf for industrial laser cutting applications. In: 2013 9th International Symposium on Mechatronics and its Applications (ISMA), pp. 1–5 (2013)
4. de Borda, J.C.: Mémoire sur les élections au scrutin. *Histoire de l'Académie Royale des Sciences* (1784)
5. Brown, L.G.: A survey of image registration techniques. *ACM Comput. Surv.* 24(4), 325–376 (1992)
6. Cabezas, I., Trujillo, M.: A method for reducing the cardinality of the pareto front. In: Alvarez, L., Mejail, M., Gomez, L., Jacobo, J. (eds.) *CIARP 2012. LNCS*, vol. 7441, pp. 829–836. Springer, Heidelberg (2012)

7. van Erp, M., Schomaker, L.: Variants of the Borda count method for combining ranked classifier hypotheses. In: Proc. of the Seventh International Workshop on Frontiers in Handwriting Recognition (7th IWFHR), pp. 443–452. iUF (2000)
8. Larson, E.C., Chandler, D.M.: Most apparent distortion: full-reference image quality assessment and the role of strategy. *Journal of Electronic Imaging* 19(1), 011006 (2010)
9. Lee, W.C., Chen, C.H.: A fast template matching method for rotation invariance using two-stage process. In: Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, IHH-MSP 2009, pp. 9–12 (2009)
10. Martin, J., Crowley, J.L.: Experimental comparison of correlation techniques. In: International Conference on Intelligent Autonomous Systems, IAS-4 (1995)
11. Nakhmani, A., Tannenbaum, A.: A new distance measure based on generalized image normalized cross-correlation for robust video tracking and image recognition. *Pattern Recogn. Lett.* 34(3), 315–321 (2013)
12. Sung-Hyuk, C.: Comprehensive survey on distance/similarity measures between probability density functions. *International Journal of Mathematical Models and Methods in Applied Sciences* 1(4), 300–307 (2007)
13. Tiirkan, M., Guillemot, C.: Image prediction: Template matching vs. sparse approximation. In: 2010 17th IEEE International Conference on Image Processing (ICIP), pp. 789–792 (2010)
14. Trujillo, M., Izquierdo, E.: A robust correlation measure for correspondence estimation. In: Proceedings of the 2nd International Symposium on 3D Data Processing, Visualization and Transmission, 3DPVT 2004, pp. 155–162 (2004)
15. Vonikakis, V., Chrysostomou, D., Kouskouridas, R., Gasteratos, A.: Improving the robustness in feature detection by local contrast enhancement. In: 2012 IEEE International Conference on Imaging Systems and Techniques (IST), pp. 158–163 (2012)
16. Vonikakis, V., Chrysostomou, D., Kouskouridas, R., Gasteratos, A.: A biologically inspired scale-space for illumination invariant feature detection. *Measurement Science and Technology* 24(7), 074024 (2013)