A robust measure for visual correspondence

Federico Tombari Luigi Di Stefano
Stefano Mattoccia
Department of Electronics Computer Science and Systems (DEIS)
Viale Risorgimento 2, 40136 - Bologna, Italy
Advanced Research Center on Electronic Systems (ARCES)
Via Toffano 2/2, 40135 - Bologna, Italy
University of Bologna
{ftombari, ldistefano, smattoccia}@deis.unibo.it

Abstract

In this paper a novel measure for visual correspondence is proposed, to be adopted for common computer vision tasks such as pattern matching, stereo vision, change detection. The proposed measure implicitly exploits the concept of order-preservation between neighbouring pixels and it is suitable for those cases where disturbance factors such as photometric distortions and occlusions occur between the images to be matched. Furthermore, this measure tends to be robust in presence of significant amount of noise which can be introduced, e.g., by cheap camera sensors. Experimental results demonstrate the effectiveness of the proposed approach in a typical template matching scenario as well as in a particular application dealing with secure gate access control.

1. Introduction

The aim of this work is to propose a novel matching measure which is robust with regards to typical disturbance factors found in real application tasks. Robust matching measures are very important in many basic computer vision tasks such as change detection, pattern matching, image registration, stereo vision. The disturbance factors which typically occur in these tasks are photometric distortions, noise and occlusions.

For what means photometric distortions between two images, they can be due to factors extrinsic to the camera, i.e. induced by variations in the light illuminating the scene or by non-lambertian surfaces viewed at different angles, as well as intrinsic, i.e. due to dynamic variations of camera parameters (e.g. auto-exposure, auto-gain) or also changes of the camera response function (e.g. when different cameras are employed). 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 in image 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, distortions 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.

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

$$NCC(I_r, I_t) = \frac{I_r \circ I_t}{||Ir|| \cdot ||I_t||} \tag{1}$$

$$ZNCC(I_r, I_t) = \frac{(I_r - \bar{I}_r) \circ (I_t - \bar{I}_t)}{||Ir - \bar{I}_r|| \cdot ||I_t - \bar{I}_t||} \tag{2}$$

with $\circ$ being the dot product, $||\cdot||$ the L_2 norm, $\bar{\cdot}$ the mean value over the patch. The most popular distance-based measures are the *Sum of Absolute Differences* (SAD) and the *Sum of Squared Differences* (SSD):

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

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

with $|\cdot|$ being the L_1 norm. Thanks to normalization with regards to the module of the vectors and to the mean intensity value of the image patch, NCC and ZNCC exhibit good

14th International Conference on Image Analysis and Processing (ICIAP 2007)
0-7695-2877-5/07 $25.00 © 2007 IEEE

robustness with respect to brightness changes, on the other hand SSD and SAD showed good insensibility towards noise [1], [11]. In addition to these measures, many alternatives have been proposed in literature with the specific aim of deploying robust image matching towards one or more of these factors: e.g. [4], [16], [17], [1], [11], [14], [7], [8]. Furthermore, many approaches have been proposed based on the so-called order-consistency or order-preservation hypothesis, that is the presence of these distortions does not violate the ordering between neighbouring pixels. These measures are called *ordinal* and some recent proposals in the change detection field are [9], [18], [13] while for what means pattern matching and stereo vision are [19], [2].

In this paper we propose a novel measure which implicitly exploits the order consistency. The proposed measure is effective in performing robust matching under real disturbance factors and can also be implemented efficiently. The description of the proposed approach is shown in Section 2, together with some further remarks contained in Section 3. Then, in Section 4 we propose some preliminary results by comparing our approach with the traditional correlation-based and distance-based measures. Finally, we draw conclusions in Section 5.

2. The proposed matching measure

The proposed approach aims to determine a measure of how well the order is preserved between corresponding pairs of neighbouring pixels in the two images. In fact, in presence of disturbance factors such as heavy photometric distortions any approach based on the absolute intensity values of the pixel would be ineffective. A simple and effective approach for evaluating the order-consistency is to consider the difference between the intensities of pairs of neighbouring pixels.

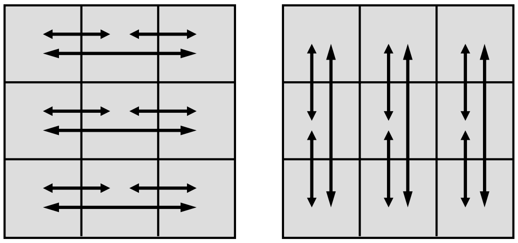

Figure 1. Considered subset of horizontal (left) and vertical (right) pairs of neighbouring pixels in a 3×3 window

To this purpose, we consider a 3×3 window centred on the pixel of the reference image patch at coordinate $(\tilde{x}, \tilde{y})$, i.e. $I_r(\tilde{x}, \tilde{y})$. In order to evaluate the order preservation between neighbouring elements within this window, many pairs (72) should be considered, as each of the 9 pixels has to be put in correspondence with each of the remaining 8. In order to simplify the problem from the computational point of view, 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 $I_r(\tilde{x}, \tilde{y})$ and $I_t(\tilde{x}, \tilde{y})$ 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. Hence, by defining the horizontal and vertical *left differences* at $I_r(\tilde{x}, \tilde{y})$ respectively as

$$R_l^h(\tilde{x}, \tilde{y}) = I_r(\tilde{x} - 1, \tilde{y}) - I_r(\tilde{x}, \tilde{y}) \tag{5}$$

$$R_l^v(\tilde{x}, \tilde{y}) = I_r(\tilde{x}, \tilde{y} - 1) - I_r(\tilde{x}, \tilde{y}) \tag{6}$$

and the horizontal and vertical *central differences* at $I_r(\tilde{x}, \tilde{y})$ as

$$R_c^h(\tilde{x}, \tilde{y}) = I_r(\tilde{x} - 1, \tilde{y}) - I_r(\tilde{x} + 1, \tilde{y}) \tag{7}$$

$$R_c^v(\tilde{x}, \tilde{y}) = I_r(\tilde{x}, \tilde{y} - 1) - I_r(\tilde{x}, \tilde{y} + 1) \tag{8}$$

and analogously $T_l^h, T_l^v, T_c^h, T_c^v$, i.e. the left and central differences at $I_t(\tilde{x}, \tilde{y})$, the proposed pointwise matching measure between pixels $I_r(\tilde{x}, \tilde{y})$ and $I_t(\tilde{x}, \tilde{y})$ can be defined as follows:

$$\begin{aligned} \varphi(\tilde{x}, \tilde{y}) = & \sum_{i=0}^{1} \sum_{j=-1}^{1} R_l^h(\tilde{x}+i, \tilde{y}+j) \cdot T_l^h(\tilde{x}+i, \tilde{y}+j) + \\ & \sum_{i=0}^{1} \sum_{j=-1}^{1} R_l^v(\tilde{x}+j, \tilde{y}+i) \cdot T_l^v(\tilde{x}+j, \tilde{y}+i) + \\ & \sum_{i=-1}^{1} R_c^h(\tilde{x}, \tilde{y}+i) \cdot T_c^h(\tilde{x}, \tilde{y}+i) + \\ & \sum_{i=-1}^{1} R_c^v(\tilde{x}+i, \tilde{y}) \cdot T_c^v(\tilde{x}+i, \tilde{y}) \end{aligned} \tag{9}$$

Typically, computer vision tasks such as pattern matching, image registration, block-based change detection, area-based stereo vision deal with the computation of a matching measure computed over an $M \times N$ *correlation area*. Hence, the straightforward extension of the proposed approach would be to compute and accumulate $\varphi(\tilde{x}, \tilde{y})$ for each element belonging to the correlation area. Nevertheless, it is easy to note that when shifting the 3×3 window over adjacent elements some correspondent neighbouring

14th International Conference on Image Analysis and Processing (ICIAP 2007)
0-7695-2877-5/07 $25.00 © 2007 IEEE

IEEE COMPUTER SOCIETY

pairs are in common. In this particular situation, it is sufficient to compute the central and the left differences for each pixel belonging to the correlation area in order to account for each neighbouring correspondence only once in the total score. This means that for each element of the correlation window only 4 pointwise correlation operations are needed, as shown in Fig. 2. Hence, the proposed measure which matches an $M \times N$ area starting at position (x_1, y_1) in the reference image vector I_r with an equally sized area starting at position (x_2, y_2) in the target image vector I_t can be seen as the sum of 4 correlation terms:

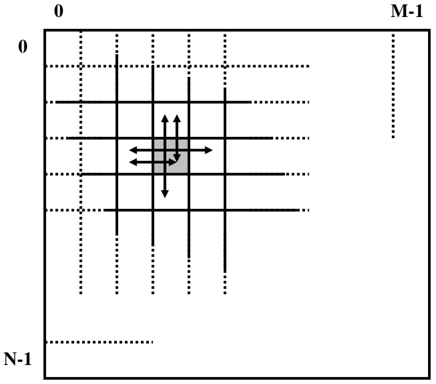

Figure 2. Considered pairs at each element of an $M \times N$ correlation area

$$\Phi(x_1, y_1, x_2, y_2) = \psi_l^h(x_1, y_1, x_2, y_2) + \psi_l^v(x_1, y_1, x_2, y_2) + \psi_c^h(x_1, y_1, x_2, y_2) + \psi_c^v(x_1, y_1, x_2, y_2) \quad (10)$$

with the generical correlation term $\psi_s^d, d \in \{v, h\}, s \in \{l, c\}$ given by:

$$\psi_s^d(x_1, y_1, x_2, y_2) = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} R_s^d(x_1 + i, y_1 + j) \cdot T_s^d(x_2 + i, y_2 + j) \quad (11)$$

Finally, it is useful to apply a normalization step currently adopted for correlation-based and other kind of measures, that is to divide $\Phi(x_1, y_1, x_2, y_2)$ by the norms of the terms involved in its computation:

$$\Phi_N(x_1, y_1, x_2, y_2) = \frac{\Phi(x_1, y_1, x_2, y_2)}{||R(x_1, y_1)|| \cdot ||T(x_2, y_2)||} \quad (12)$$

where

$$||R(x_1, y_1)|| = \sqrt{\sum_{d \in \{v,h\}} \sum_{s \in \{l,c\}} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \left(R_s^d(x_1, y_1)\right)^2} \quad (13)$$

$$||T(x_2, y_2)|| = \sqrt{\sum_{d \in \{v,h\}} \sum_{s \in \{l,c\}} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \left(T_s^d(x_2, y_2)\right)^2} \quad (14)$$

This allows for more robustness towards spatially constant multiplicative brightness distortions, and also renders $\Phi_N(x_1, y_1, x_2, y_2)$ belonging to the interval $[-1, +1]$. It is also important to note that in many scenarios where Φ_N has to be computed on several neighbouring image patch vectors (e.g. pattern matching, stereo vision) the calculation of the terms at denominator of (12) can be efficiently performed by means of incremental techniques [12], [5], hence this normalization step adds only a small - often neglectible - computational overhead to the overall process.

The proposed matching measure, Φ_N, can be easily integrated in any matching task where several $M \times N$ candidates are extracted from a *search area* and matched with a $M \times N$ pattern, with the aim of selecting the candidate most *similar* to the pattern, i.e. the best matching candidate. For instance, this is a recurring task for pattern matching and image registration scenarios. In order to deploy this, given a set of coordinates $(x_u, y_u) \in U$ denoting the candidates and a pair of coordinates $(\tilde{x}, \tilde{y})$ denoting the pattern, Φ_N is computed between the pattern and each candidate and the best matching coordinates (x_w, y_w) are selected as those corresponding to the candidate which maximizes the matching measure:

$$(x_w, y_w) = \arg\max_{u \in U} \{\Phi_N(\tilde{x}, \tilde{y}, x_u, y_u)\} \quad (15)$$

3 Further remarks

In order to give a behavioural interpretation of the proposed matching measure Φ_N, it can be noted that the pointwise score depends only from the local intensity variations around each point, which are proportional to the local informative content of the image. In particular, edge points tend to generate important correlation coefficients. Conversely, uniform areas does not improve - nor decrease - the correlation score due to the fact the local differences are close to zero. Generally speaking, even heavy non-linear real photometric distortions tend not to invert the order of the majority of correspondent neighbouring pixels in textured areas (otherwise the informative content would be totally lost) and thus matching a pattern under these circumstances yields a positive correlation score. Conversely, non-matching patterns tend to yield very low - or even negative - scores as the order of the neighbouring pixels tend to be randomly violated. Analogous considerations apply in presence of noise and occlusions.

Finally, from a computational point of view, as it can be noted from (10) the computation of Φ_N requires the calculation of 4 distinct correlation terms. Hence, if lateral and central differences are computed at start-up and stored, the additional calculations required for a pattern matching process based on the proposed approach can be estimated as 4 times those needed by a conventional approach based on cross-correlation. Furthermore, it is important to note that any fast exhaustive [10], [6] or non-exhaustive [3], [15] cross-correlation based methods could be adopted in order to further reduce the overall computational load.

4 Experimental results

This section presents preliminary experimental results aimed at comparing the effectiveness of the proposed measure with respect to the traditional approaches in some typical scenarios where images are affected by real disturbance factors. The experimental section is divided into 3 experiments. Experiment 1 and 2 deal with two typical template matching scenarios while Experiment 3 shows some preliminary results dealing with a change detection process applied to monitor a high security gate.

4.1 Experiment 1

In Experiment 1, a typical template matching task is proposed: 2 templates are extracted from an image taken with a digital camera with a 3 Mega Pixels (MP) sensor and under good illumination conditions. The templates have to be matched with 4 images taken successively with a different, cheaper sensor (mobile phone camera with 1.3 MP sensor), as shown in Fig. 3. Photometric disturbance factors are present, as the two cameras are in slightly different positions and real illumination changes are applied by means of moving the rheostat of a lamp ($L1$), by using a torch light in the center of the scene($L2$, $L3$) and by using the camera flash ($L4$). The use of different camera sensors also induce photometric distortions due to different intrinsic camera parameters under different illumination conditions. Noise is introduced because of the use of different camera sensors and because of the different illumination conditions. All images have been converted to 8-bit grayscale images after acquisition as colour images.

The proposed measure, Φ_N, is tested against the traditional matching measures NCC, ZNCC and SSD. Table 1 shows the results of the comparison, where *Yes* indicates that the resulting matching position is the correct one while *No* identifies an incorrect match. As it can be seen, the proposed measure is the best performing one, leading to 8 correct matches out of 8 tests. Between the other approaches, ZNCC is the measure which performs best ($6/8$), while SSD and NCC are not suitable to work under these kind of distortions.

Table 1. Comparison of different measures, Experiment 1

T	I	NCC	ZNCC	SSD	NEW
T1	L1	No	Yes	No	Yes
T1	L2	No	No	No	Yes
T1	L3	No	Yes	No	Yes
T1	L4	No	Yes	No	Yes
T2	L1	No	Yes	No	Yes
T2	L2	No	Yes	No	Yes
T2	L3	No	No	No	Yes
T2	L4	Yes	Yes	No	Yes
		1/8	6/8	0/8	8/8

4.2 Experiment 2

In Experiment 2, another template matching task is proposed which unlike the previous one deals with the presence of occlusions. A template is extracted from an image taken with a 1.3 MP mobile phone camera sensor under good illumination conditions, and it is matched with 8 images taken successively with a 0.3 MP mobile phone camera sensor under different illumination condition, as shown in Fig. 4. Along the 8 images different occlusions occur to the object to be matched. As before, noise is introduced by the use of different, cheap camera sensors under different illumination conditions. All images have been converted to 8-bit grayscale images after acquisition as colour images.

Similarly to Experiment 1, a comparison is shown between the proposed measure, Φ_N, and the matching measures NCC, ZNCC and SSD. Table 2 shows the results of the comparison. Again, the proposed measure is the one which performs best, with 7 correct matches out of 8 trials. Unlike to Experiment 1, in this case ZNCC performs poorly and its score is closer to those of SSD and NCC, confirming that ZNCC is more sensible to occlusions than to photometric distortions.

Table 2. Comparison of different measures, Experiment 2

I	NCC	ZNCC	SSD	NEW
O1	No	No	No	Yes
O2	No	No	No	Yes
O3	No	Yes	No	Yes
O4	No	No	No	Yes
O5	No	Yes	No	Yes
O6	No	No	No	Yes
O7	No	No	No	Yes
O8	No	No	No	No
	0/8	2/8	0/8	7/8

4.3 Experiment 3

Experiment 3 deals with a case study where access to a high security gate has to be monitored[1]. In order to do this we apply a change detection algorithm to incoming frames in order to detect foreground regions by comparison against a background image (taken when the gate is empty). The change detection algorithm relies on the proposed measure: at each point of the current frame and background image Φ_N 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 light coming from outside.

Fig. 5 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.

5 Conclusions and future work

A novel measure for robust visual correspondence under disturbance factors such as photometric distortions, noise and occlusions has 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 measure seems effective in retrieving the correct match in images affected by real distortions and promising in a challenging application scenario considered as a case study. Future work is focused on the generalization of the concept on which the proposed approach is based to a novel class of matching measures, and on the comparison of the proposed measure with more recent proposals.

References

[1] P. Aschwanden and W. Guggenbuhl. 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, 1992.

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

[3] K. Briechle and U. Hanebeck. Template matching using fast normalized cross correlation. In *Proc. of SPIE AeroSense Symposium*, volume 4387, Orlando, Florida, USA, 2001.

[4] A. Crouzil, L. Massip-Pailhes, and S. Castan. A new correlation criterion based on gradient fields similarity. In *Proc. Int. Conf. Pattern Recognition (ICPR)*, pages 632–636, 1996.

[5] F. Crow. Summed-area tables for texture mapping. *Computer Graphics*, 18(3):207–212, 1984.

[6] L. Di Stefano, S. Mattoccia, and F. Tombari. Zncc-based template matching using bounded partial correlation. *Pattern Recognition Letters*, 26(14):2129–2134, 2005.

[7] A. J. Fitch, A. Kadyrov, W. J. Christmas, and K. J. Orientation correlation. In P. Rosin and D. Marshall, editors, *British Machine Vision Conference*, volume 1, pages 133–142, 2002.

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

[9] A. Lanza and L. Di Stefano. Detecting changes in grey level sequences by *ML* isotonic regression. In *Proc. IEEE Int. Conf. on Video and Signal Based Surveillance (AVSS)*, page 4, 2006.

[10] J. Lewis. Fast template matching. *Vision Interface*, pages 120–123, 1995.

[11] J. Martin and J. Crowley. Experimental comparison of correlation techniques. In *Proc. Int. Conf. on Intelligent Autonomous Systems*, volume 4, pages 86–93, 1995.

[12] M. Mc Donnel. Box-filtering techniques. *Computer Graphics and Image Processing*, 17:65–70, 1981.

[13] A. Mittal and V. Ramesh. An intensity-augmented ordinal measure for visual correspondence. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 849–856, 2006.

[14] F. Odone, E. Trucco, and A. Verri. General purpose matching of grey level arbitrary images. In C. Arcelli, L. Cordella, and G. Sanniti di Baja, editors, *4th Int. Workshop on Visual Form, LNCS*, pages 573–582. Springer-Verlag, 2001.

[15] A. Rosenfeld and G. Vanderburg. Coarse-fine template matching. *IEEE Trans. on Sys., Man and Cyb.*, 7:104–107, 1977.

[16] D. Scharstein. Matching images by comparing their gradient fields. In *Proc. Int. Conf. Pattern Recognition (ICPR)*, volume 1, pages 572–575, 1994.

[17] P. Seitz. Using local orientational information as image primitive for robust object recognition. In *Proc. SPIE, Visual Communication and Image Processing IV*, volume 1199, pages 1630–1639, 1989.

[18] B. Xie, V. Ramesh, and T. Boult. Sudden illumination change detection using order consistency. *Image and Vision Computing*, 22(2):117–125, 2004.

[19] R. Zabih and J. Woodfill. Non-parametric local transforms for computing visual correspondence. In *Proc. European Conf. Computer Vision*, pages 151–158, 1994.

[1] The authors would like to thank Plexa Srl. for the grant of publishing the material shown in Fig. 5

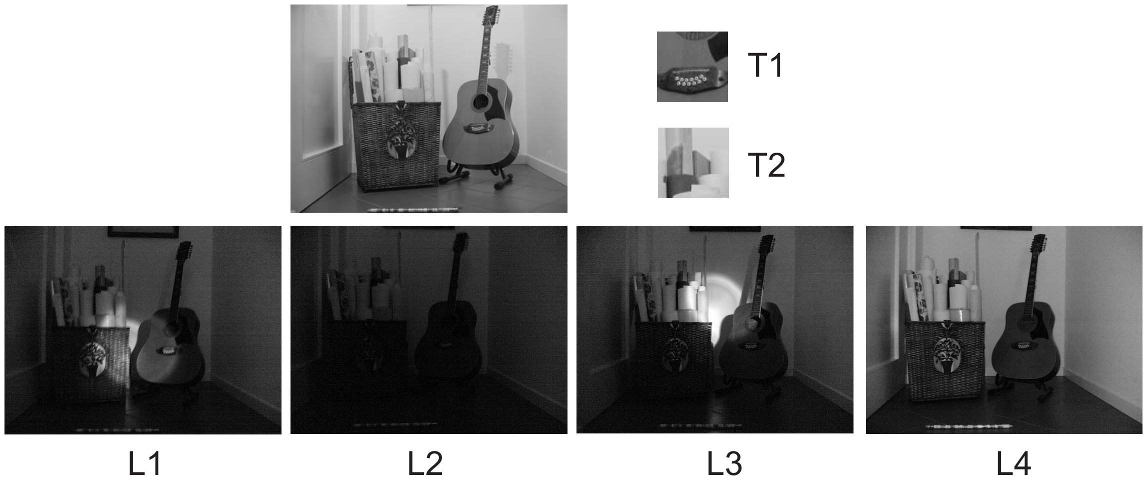

Figure 3. Templates $T1, T2$ and images $L1-L4$ used in Experiment 1

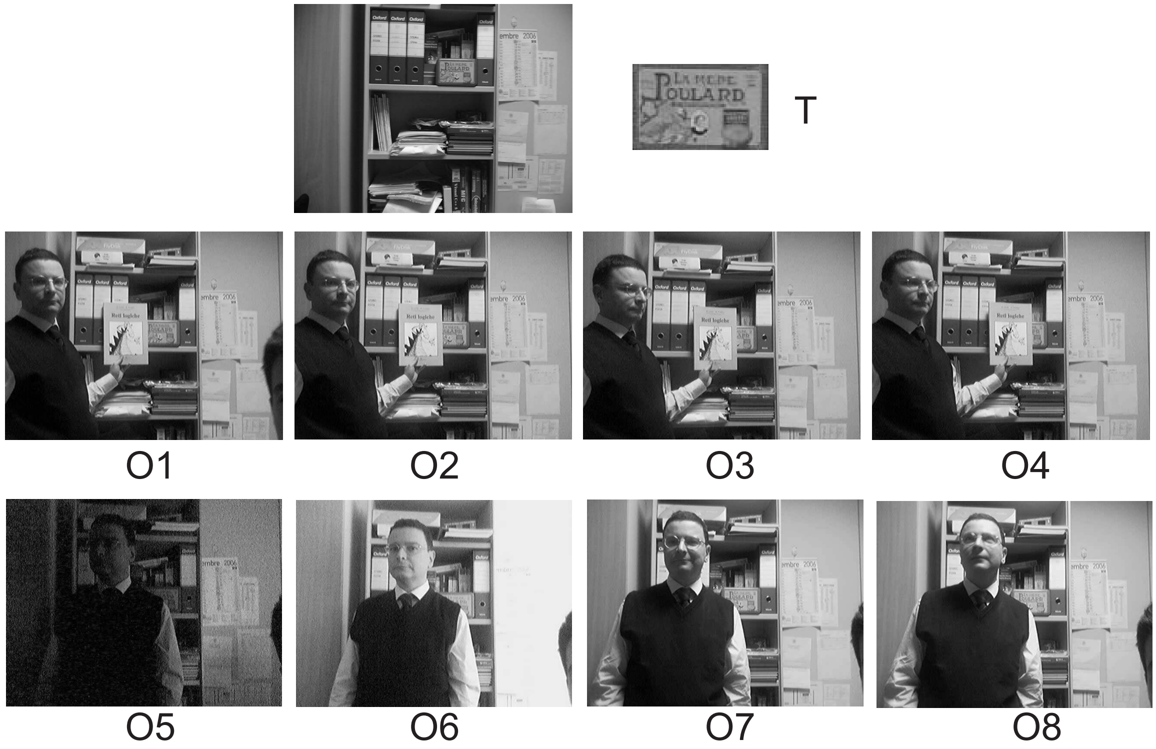

Figure 4. The template and images $O1-O8$ used in Experiment 2

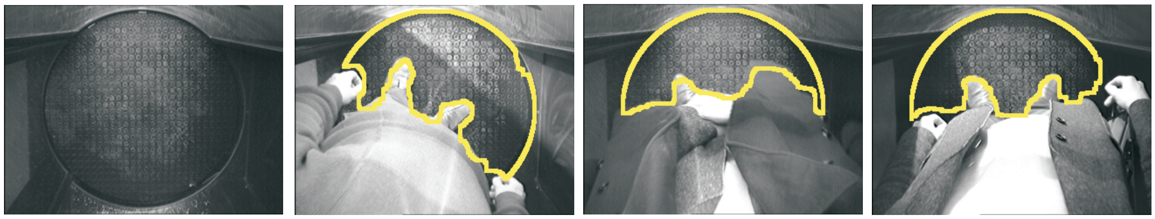

Figure 5. The background (*left*) and 3 examples dealing with the case study of Experiment 3

14th International Conference on Image Analysis and Processing (ICIAP 2007)
0-7695-2877-5/07 $25.00 © 2007 **IEEE**