EFFICIENT OBJECT DETECTION ROBUST TO RST WITH MINIMAL SET OF EXAMPLES

Sebastien Onis, Henri Sanson, Christophe Garcia
France Telecom RD, 4 rue du clos courtel, Rennes, France

Jean-Luc Dugelay
Eurecom, Sophia Antipolis, France

Keywords: Object detection, correlation, affine deformation.

Abstract: In this paper, we present an object detection approach based on a similarity measure combining cross-correlation and affine deformation. Current object detection systems provide good results, at the expense of requiring a large training database. The use of correlation anables object detection with very small training set but is not robust to the luminosity change and RST (Rotation, Scale, translation) transformation. This paper presents a detection system that first searches the likely positions and scales of the object using image preprocessing and cross-correlation method and secondly, uses a similarity measure based on affine deformation to confirm or not the predetection. We apply our system to face detection and show the improvement in results due to the images preprocessing and the affine deformation.

1 INTRODUCTION

Object detection is a classical research topic. Most of the current object detection systems use machine learning like Gaussian Mixture Model, Neural Networks or Support Vector Machine. In (Viola and Jones, 2001) the system performs fast object detection using a cascade of classifiers associated with Haar descriptors. In (Santiago-Mozos et al., 1999) the detection system extracts features using PCA and a classifier based on SVM method to detect objects in infrared images. (Garcia and Delakis, 2004) perform face detections using a convolutional neural network and in (Sung and Poggio, 1998) face detection is done using GMM to extract face descriptors and a perceptron to perform classification. These systems currently provide the best detection rate, however the features used are dependent on the object to detect. Additionally, they need a large training database, manually annotated to initialize the detection system, which represents long and tiresome work. Thus for each object to detect, it is necessary to choose or learn good features and to build a training database.

Correlation is a well-known shape detection method which has many advantages; easy to implement, fast, easily adapable to a broad variety of shapes and not requiring complex feature extractors, or a large training database. This method however, is not robust to illumination change, scale variations or rotation.

We describe in this paper an object detection system based on cross-correlation, robust to illumination changes and affine deformations. (MacLean and Tsotsos, 2007) performs shape detection, using normalized cross-correlation for various object scales using a pyramid of images. The use of deformation models for object detection produced interesting results. (Edwards et al., 1999) performs face detection using Active Appearance Model deforming the faces textures in order to maximise the similarity between the images to compare. (Wakahara et al., 2001) shows that affine deformation increases the robustness in rotation and scale changes of a character recognition system based on cross-correlation measures. Our system performs a predetection using normalized cross-correlation on a pyramid of images. We then use similarity measure based on affine deformation and centered normalized cross-correlation to valid or not the predetection.

Section II describes the predetection system based on the normalized cross-correlation applied to a filtered pyramid of images. Section III is about the de-

Onis S., Sanson H., Garcia C. and Dugelay J. (2008).
EFFICIENT OBJECT DETECTION ROBUST TO RST WITH MINIMAL SET OF EXAMPLES.
In *Proceedings of the Third International Conference on Computer Vision Theory and Applications*, pages 179-185
DOI: 10.5220/0001083601790185

cision process which consists of determining whether a predetection is valid or not. Finally in section IV we apply our system to face detection and analyse the influence of the image filters and the affine deformation compensation upon the detection rate.

2 PREDETECTION

The first step of our system consists of detecting the likely positions and scales of the searched object. The system is based on the normalized cross-correlation between each example image of the object to detect and a pyramid of filtered images.

2.1 Normalized Cross-Correlation

This section introduces the well-known normalized cross-correlation method used for object predetection. We denote the reference image F and the test image G. We represent F and G by grey level functions $f(\mathbf{r})$ and $g(\mathbf{r})$. $\mathbf{r}$ denotes a 2D loci vector (u,v).

An object is predetected at position $\mathbf{p} = (i,j)$ in G if this point is a local maximum of the normalized cross-correlation function $C(\mathbf{p})$ and if this maximum is greater than a given threshold.

$$\sigma_f = \sqrt{\sum_{\mathbf{r}\in Dom_F} f(\mathbf{r})^2}$$

$$\sigma_g = \sqrt{\sum_{\mathbf{r}\in Dom_F} g(\mathbf{p}+\mathbf{r})^2}$$

$$C(\mathbf{p}) = \frac{1}{\sigma_f \sigma_g} \sum_{\mathbf{r}\in Dom_F} f(\mathbf{r}) g(\mathbf{p}+\mathbf{r}) \qquad (1)$$

Interestingly enough, we can easily show that the similarity measure based on the normalized cross-correlation and the $L2$ distance between two normalized images F' and G' respectively represented by the grey level functions $\frac{f(\mathbf{r})}{\sigma_f}$ and $\frac{g(\mathbf{r})}{\sigma_g}$ are equivalent. Indeed, if $D(\mathbf{p})$ is the L2 distance between the images G' and F' at position $\mathbf{p}$ in G'

$$\begin{aligned} D(\mathbf{p}) &= \sum_{\mathbf{r}\in Dom_F} \left(\frac{f(\mathbf{r})}{\sigma_f} - \frac{g(\mathbf{p}+\mathbf{r})}{\sigma_g} \right)^2 \\ &= \overbrace{\sum_{\mathbf{r}\in Dom_F} \left(\frac{f(\mathbf{r})}{\sigma_f} \right)^2}^{=1} + \overbrace{\sum_{\mathbf{r}\in Dom_F} \left(\frac{g(\mathbf{p}+\mathbf{r})}{\sigma_g} \right)^2}^{=1} \\ &\quad - \frac{2}{\sigma_f \sigma_g} \sum_{\mathbf{r}\in Dom_F} f(\mathbf{r}) g(\mathbf{p}+\mathbf{r}) \qquad (2) \end{aligned}$$

Then $D(\mathbf{p}) = 2(1 - C(\mathbf{p}))$ only depends on the normalized cross-correlation.

2.2 Image Processing for Predetection

The defined similarity measure applied to grey-scale images gives results of poor precision (Fig. 5). In order to increase the robustness of the predetection system to illumination variations, we apply a high pass filter inspired from the Nagano method to the images F and G. This filter extracts the edges of the images. Thus, the predetection system becomes a measure of the edges similarity. If all the edges are perfectly superposed, the normalized cross-correlation score is 1 and the less the edges are superposed, the closer to 0 the similarity score approaches.

v_1, v_2, v_3 and v_4 corresponding to the 4 following matrix:

$$\begin{pmatrix} 1 & 1 & 0 & -1 & -1 \\ 1 & 1 & 0 & -1 & -1 \\ 1 & 1 & 0 & -1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \\ -1 & -1 & -1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & -1 & -1 & -1 \\ 0 & 0 & 0 & -1 & -1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & -1 & 0 & 1 & 0 \\ -1 & -1 & -1 & 0 & 0 \\ -1 & -1 & 0 & 0 & 0 \end{pmatrix}$$

$$f_1 = \max\left(|f \otimes v_1|, |f \otimes v_2|, |f \otimes v_3|, |f \otimes v_4|\right)$$

Filter f_1 convolutes the image F represented by the function $f(\mathbf{r})$ with 4 filters represented by the matrix (v_1, v_2, v_3, v_4). Each filter is an edge detector in a given direction. The final image represented by the function $f_1(\mathbf{r})$ is the maximum of the four edges values of F detected using the filters (v_1, v_2, v_3, v_4).

Figure 1: Predetection filter applied to a face image.

2.3 Implementation Using the Pyramid of Images

In order to create a predetection system able to detect objects of different sizes, the test images are repeatedly down-sampled by a factor of 1.2, resulting in a pyramid of images (Fig. 2). Each image of the pyramid is filtered using the predetection filter. Then we apply the normalized cross-correlation detection method to each image of the pyramid and each filtered reference image. The predetection system searchs the likely positions and scales of the researched object with recall close to 1 (number of good detections divided by the number of elements to detect). The next step consists of refining and verifying the predetection information in order to increase the precision of

the system (number of good detections divided by the number of detections).

Figure 2: Example of a Pyramid of images used for predetection. According to the image of the pyramid where an object is predetected, we know the dimension of the predetected object compared to the reference one.

3 DECISION

Once the predetection phase has been carried out, we apply the decision system to each predetected object. The decision system uses a similarity measure based on centered normalized cross-correlation and affine deformation compensation. Each predetected object is deformed in order to maximize the similarity between the deformed test image and the corresponding reference image of the object. Then, the images of the objects to compare are preprocessed via a histogram equalization, a high pass filter and image normalizations. Finally we apply the centered cross-correlation to obtain a similarity score between the two images deformed and preprocessed to compare.

3.1 Affine Deformation Compensation

This section describes the computational model used for optimal affine deformation determination. The key idea is to find the maximum similarity measure for the affine deformation parameters. We first describe the chosen similarity measure and the corresponding function Ψ to maximize. We then explain the Gauss Newton optimization method used to find a maximum of function Ψ.

3.1.1 Formulation of the Affine Deformation Method

Affine deformation is the first-order approximation of the image deformation resulting from the perspective projection of a rigid plane object which undergoes a displacement and a rotation. Affine deformation consists in translating, tilting and changing the vertical and horizontal scale of an image.

If $G^* = \{g^*(\mathbf{r})\}$ is the result of an affine transformation of a grey-scale image $G = \{g(\mathbf{r})\}$. The coordinates $(0,0)$ being the image centre, we can write:

$$g^*(\mathbf{r}) = g(\mathbf{r} + \mathbf{d_r})$$

$$\mathbf{r} = \begin{pmatrix} u \\ v \end{pmatrix}$$

$$\mathbf{d_r} = \begin{pmatrix} d_u \\ d_v \end{pmatrix} = \begin{pmatrix} a_0 u + a_1 v + a_2 \\ a_3 u + a_4 v + a_5 \end{pmatrix}$$

The 6 parameters $(a_0, ..., a_5)$ define the affine deformation. a_2 and a_5 are the translation parameters, a_0, a_1, a_2 and a_3 determine the image tilt and scale.

The criterion usually used to determine the best affine deformation is the minimization of the $L2$ distance between the images requiring matching. In order to ensure robustness versus illumination, we introduce here the criterion of maximizing the centered normalized cross-correlation of the deformed reference image and the test image, namely: find the parameters $(a_0, ..., a_5)$ which maximize the following objective function Ψ.

$$\Psi = \sum_{\mathbf{r} \in Dom_F} \overbrace{\left(\frac{f(\mathbf{r}) - m_f}{\sigma_f} \right)}^{f_n} \underbrace{\left(\frac{g(\mathbf{p} + \mathbf{r} + \mathbf{d_r}) - m_g}{\sigma_g} \right)}_{g_n} \tag{3}$$

$F = \{f(\mathbf{r})\}$ and $G = \{g(\mathbf{r})\}$ are respectively the reference and the test image, $\mathbf{p}$ the coordinate of G where the object have been predetected.
m_f and m_g are the means of the functions $f(r)$ and $g^*(\mathbf{p} + \mathbf{r})$, $\mathbf{r} \in Dom_F$:

$$m_f = \sum_{\mathbf{r} \in Dom_F} f(\mathbf{r})$$

$$m_g = \sum_{\mathbf{r} \in Dom_F} g(\mathbf{p} + \mathbf{r} + \mathbf{d_r})$$

σ_f and σ_g are the standard deviations of the functions $f(\mathbf{r})$ and $g^*(\mathbf{p} + \mathbf{r})$, $r \in Dom_F$:

$$\sigma_f = \sqrt{\sum_{\mathbf{r} \in Dom_F} (f(\mathbf{r}) - m_f)^2}$$

$$\sigma_g = \sqrt{\sum_{\mathbf{r} \in Dom_F} (g(\mathbf{p} + \mathbf{r} + \mathbf{d_r}) - m_g)^2}$$

We notice that only the functions g, m_g and σ_g depend on the affine deformation parameters.

3.1.2 Optimal Affine Deformation Determination

We describe in this section the computational model used to determine the affine deformation parameters. First of all, following the necessary condition of Ψ maximization yields to a set of six equations

$$\frac{\partial \Psi}{\partial a_i} = 0 \qquad i \in [0,5] \tag{4}$$

These equations cannot be solved analytically. Since the problem has a low dimension, it seems appropriate to determine the affine deformation parameters using non linear optimisation. (Dugelay and Sanson, 1995) shows that the Gauss Newton iterative method enables a robust and fast convergence solution for affine deformation optimization.

This method uses two approximations to perform the optimization:

- The function Ψ to optimize is locally a second-order polynomial function.
- The second derivative of the function g is 0 (the Hessian matrix of $g(\mathbf{r})$, $H_g = \mathbf{0}$). Namely, that the luminance variation of the image G is locally linear.

We denote $A_k = \begin{pmatrix} a_0 & a_1 & a_2 & a_3 & a_4 & a_5 \end{pmatrix}^t$ the value of the affine deformation parameters to the k^{th} iteration.

Using the approximation Ψ is locally a second-order polynomial function, the updating parameter vector is given by:

$$A_{k+1} = A_k - H_A^{-1} G_A \tag{5}$$

Where H_A is the Hessian of the cost function Ψ and G_A its gradient.

$$G_A = \begin{pmatrix} \frac{\partial \Psi}{\partial a_i} \\ \vdots \end{pmatrix} \quad H_A = \begin{pmatrix} \frac{\partial^2 \Psi}{\partial a_i \partial a_j} & \cdots \\ \vdots & \ddots \end{pmatrix}$$

To simplify, henceforth we will use:

$$\begin{array}{ll} g_p \Leftrightarrow g(\mathbf{p}+\mathbf{r}+\mathbf{d_r}) & \\ \mathbf{g_r} \Leftrightarrow \nabla_r(g_p) & \mathbf{d_r^i} \Leftrightarrow \frac{\partial \mathbf{d_r}}{\partial a_i} \\ m_g^i \Leftrightarrow \frac{\partial m_g}{\partial a_i} & \sigma_g^i \Leftrightarrow \frac{\partial \sigma_g}{\partial a_i} \end{array}$$

- g_p value of g at point $(\mathbf{p}+\mathbf{r}+\mathbf{d_r})$.
- $\mathbf{g_r}$ the gradient value of g at point $(\mathbf{p}+\mathbf{r}+\mathbf{d_r})$.
- $\mathbf{d_r^i}$ the derivative function of $\mathbf{d_r}$ with respect to a_i.
- m_g^i the derivative function of the mean of $g(\mathbf{p}+\mathbf{r}+\mathbf{d_r})$, $\mathbf{r} \in Dom_F$.
- σ_g^i the derivative function of the standard deviation of $g(\mathbf{p}+\mathbf{r}+\mathbf{d_r})$, $\mathbf{r} \in Dom_F$

In order to determine A_{k+1}, we have to compute each iteration the matrix G_A and H_A. The assumption of the local linear variation of $g(\mathbf{r})$ allows us to determine G_A and H_A using only the known functions $f(\mathbf{r})$ and $g(\mathbf{p}+\mathbf{r}+\mathbf{d_r})$, the gradient of $g(\mathbf{r})$ (easily computable using bilinear approximation), and $\mathbf{d_r^i}$.

Indeed, G_A is given by:

$$\begin{aligned} \frac{\partial \Psi}{\partial a_i} &= \sum_{\mathbf{r} \in Dom_F} \left(\frac{f(\mathbf{r}) - m_f}{\sigma_f} \right) \frac{\partial}{\partial a_i} \left(\frac{g_p - m_g}{\sigma_g} \right) \\ &= \sum_{\mathbf{r} \in Dom_F} f_n \frac{\partial g_n}{\partial a_i} \qquad (6) \\ &= \sum_{\mathbf{r} \in Dom_F} f_n \frac{\left(\mathbf{d_r^i}^t \mathbf{g_r} - m_g^i \right) \sigma_g - (g_p - m_g) \sigma_g^i}{\sigma_g^2} \end{aligned}$$

With :

$$m_g^i = \sum_{\mathbf{r} \in Dom_F} \mathbf{d_r^i}^t \mathbf{g_r} \tag{7}$$

If we denote $V_g^i = \frac{\partial V_g}{\partial a_i}$ the derivative function of the variance V_g of $g(\mathbf{p}+\mathbf{r}+\mathbf{d_r})$:

$$\begin{aligned} \sigma_g^2 = \quad V_g &= \sum_{\mathbf{r} \in Dom_F} (g_p - m_g)^2 \\ V_g^i &= \sum_{\mathbf{r} \in Dom_F} 2 \left(\mathbf{d_r^i}^t \mathbf{g_r} - m_g^i \right) (g_p - m_g) \\ \sigma_g^i &= \frac{V_g^i}{2 V_g^{\frac{1}{2}}} \qquad (8) \end{aligned}$$

Similarly, noticing that $\forall (i,j), \frac{\partial \mathbf{d_r^i}}{\partial a_j} = \mathbf{0}$, the Hessian matrix H_A is determined as follows:

$$\frac{\partial^2 \Psi}{\partial a_i \partial a_j} = \sum_{\mathbf{r} \in Dom_F} f_n \frac{\partial^2 g_n}{\partial a_i \partial a_j} \tag{9}$$

$$\begin{aligned} \sigma_g^4 \frac{\partial^2 g_n}{\partial a_i \partial a_j} = \qquad & (10) \\ & 2 (g_p - m_g) \sigma_g \sigma_g^i \sigma_g^j - \left(\mathbf{d_r^j}^t \mathbf{g_r} - m_g^j \right) \sigma_g^i \sigma_g^2 - \\ & (g_p - m_g) \sigma_g^{ij} \sigma_g^2 - \left(\mathbf{d_r^i}^t \mathbf{g_r} - m_g^i \right) \sigma_g^j \sigma_g^2 \end{aligned}$$

$\sigma_g^{ij} = \frac{\partial^2 \sigma_g}{\partial a_i \partial a_j}$ is the second derivative function of σ_g.

$$\begin{aligned} \sigma_g^{ij} &= \frac{\partial}{\partial a_j} \frac{1}{2} \left(V_g^i V_g^{\frac{-1}{2}} \right) \\ &= \frac{1}{2 V_g} \left(V_g^{ij} \sigma_g - \frac{V_g^i V_g^j}{2 \sigma_g} \right) \qquad (11) \end{aligned}$$

With $V_g^{ij} = \frac{\partial V_g^i}{\partial a_j}$ the second derivative function of the variance V_g:

$$V_g^{ij} = \sum_{\mathbf{r} \in Dom_F} 2 \left(\mathbf{d_r^i}^t \mathbf{g_r} - m_g^i \right) \left(\mathbf{d_r^j}^t \mathbf{g_r} - m_g^j \right) \tag{12}$$

After the computation of G_A and H_A, for the k^{th} iteration, we compute H_A^{-1}. In practice this inversion

does not raise any problems. Finally using (5), the affine deformation system converges towards solution in less than 10 iterations.

3.2 Image Preprocessing for Decision

In order to reduce the sensitivity of the decision system to variations of illumination, we apply the following image preprocessing to the deformed object images to compare.

The image preprocessing is performed in 3 steps:

- Histogram equalization:
  Histogram equalization is a contrast enhancement technique with the objective to obtain a new image with uniform histogram. This method usually increases the local contrast of an image, and reduces the variability of the grey-scale images representing the object we have to detect.
- High Pass Filter:
  Image low frequency information are usually not pertinent for the detection using cross-correlation, that is why we substract from both images to compare their corresponding blurred images. If we denote $G_1 = g_1(\mathbf{r})$ the image $G = g(\mathbf{r})$ filtered by the high pass filter.

$$g_1(\mathbf{r}) = g(\mathbf{r}) - Blur\,(g(\mathbf{r}))$$
$$Blur\,(g(\mathbf{r})) = g(\mathbf{r}) \otimes w(\mathbf{r},n)$$

  With $w(\mathbf{r},n) = \frac{1}{4n}$ if $\|\mathbf{r}\|_\infty < n$ else $w(\mathbf{r},n) = 0$.
- Sigmoid normalization:
  The sigmoid normalization maximises the low gray scale values, minimises the high ones and thus standardizes the distribution of grey scale values of the image, thus increasing the precision of our detection system (Fig. 5). If $G_2 = g_2(\mathbf{r})$ is the normalized image, then:

$$g_2(\mathbf{r}) = Sig\,(g_1(\mathbf{r}))$$
$$Sig(x) = 1 - \frac{2}{1+e^{-ax}}$$

  The value of a is about 20 in our detection system.

Figure 3: Decision Preprocessing applied to a face image. From the left to the right, grey-scale face image, histogram equalization, high pass filter and finally, sigmoid normalization.

4 EXPERIMENTAL RESULTS

In this section, we first present results that confirm robustness in rotation and scale changes of the similarity measure based on affine deformation compensation and normalized centered cross-correlation. Then we apply the detection system to faces, using a test database containing 450 faces and show the improvement brought by the proposed method.

4.1 Affine Deformation Evaluation

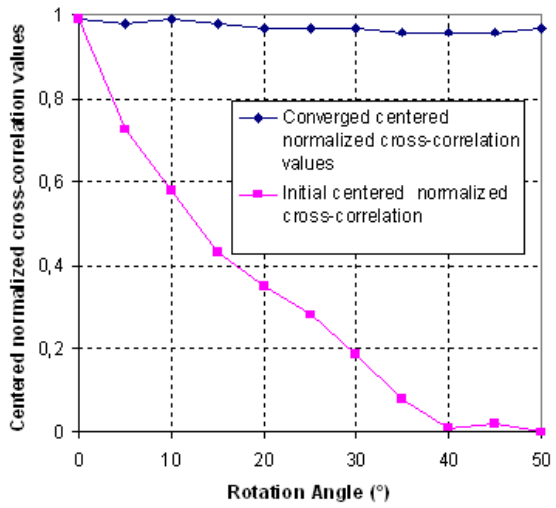
The purpose of the affine deformation compensation is to bring robustness versus rotation, scale changes and translation to the centered normalized cross-correlation similarity measure. This section shows two quantitative results obtained by applying our affine deformation method to a 35×41 pixel face image with a wide variety of pure rotation, and scale change.

Fig. 4(a) shows centered normalized cross-correlation score between an input grey-scale face image and the corresponding artificially generated image applying pure rotation. It is clear that until a rotation of about 50°, the affine deformation method converges and the similarity measure is almost invariant to rotation.

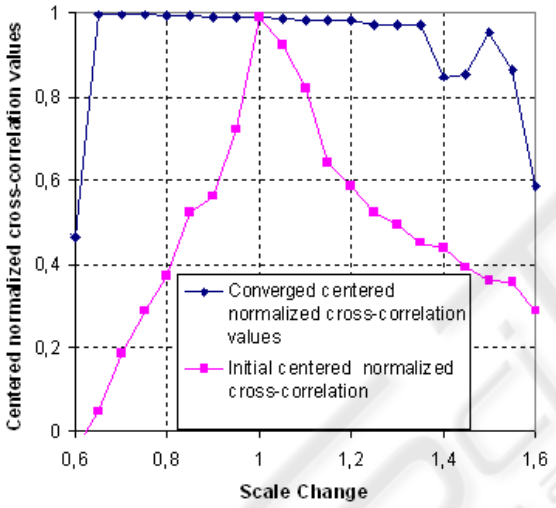
We reproduce the same experiment applying pure scale change to the artificially generated image. We can see on Fig. 4(b) that if the affine deformation converges to the optimal solution, the centered normalized cross-correlation value is about 1. The values of the converged centered normalized cross-correlation lower than 0.9 are due to local maximum convergence of the affine deformation optimization algorithm.

4.2 Detection Evaluation

In order to evaluate our system, we apply it to face detection using a test base containing 450 faces. The reference database consists of 15 faces Fig. (6), selected in order to obtain a good representation of the faces space with a minimal set of examples. Fig. 5 shows the relation between the precision (number of good detections divided by the number of detections) and the recall (number of good detections divided by the number of elements to detect). Thus, the better a detector is, the closer the corresponding roc-curve is to the upper right corner. We notice the predetection system is able to detect most of the test database faces but with poor precision. The decision system using centered normalized cross-correlation on grey scale images clearly increases the detection precision. We notice the relevance of the decision system images



(a) Relation between the mean normalized cross-correlation values and the rotation.



(b) Relation between the mean normalized cross-correlation values and scale change.

Figure 4: Affine deformation experimental results.

preprocessing and the affine deformation. The precision of our detection system for a recall of 0.9 without image preprocessing and affine deformation compensation is 0.28, the image preprocessing increases the precision to 0.55 and the affine deformation to 0.79.

This system introduces promising methods to perform efficient detection with very small training set. However, it should be noted that we are not able to obtain good detection rates from complex face detection databases like CMU , where lots of faces are occluded and very badly contrasted. Our future works is to produce a detection system using reduced training sets able to reach detection rates close to state-of-the-art.

5 CONCLUSIONS

The object detection system based on the cross-correlation method is sensitive to illumination changes, rotations, translations and scale changes. To solve this problem, we introduce a detection process divided in a predetection and a decision system. The two parts of the detection system use different image preprocessing which increases the detection speed and rates. This method has shown good results on face detection. Additionally, we introduce a new similarity measure based on cross-correlation and affine deformation. The affine deformation system based on the mean normalized cross-correlation optimization we have developed is very promising, and shows good convergence for complex grey-scale images. Thus the measure we use for decision is robust to RST and increases the precision of our detection system.

REFERENCES

Dugelay, J. L. and Sanson, H. (1995). Differential methods for the identification of 2d and 3d motion models in image sequences. In *Signal Processing: Image Communication 7*.

Edwards, G. J., Cootes, T. F., and Taylor, C. J. (1999). Advances in active appearance models. In *Computer Vision, 1999*. INSTICC Press.

Garcia, C. and Delakis, M. (2004). Convolutional face finder, a neural architecture for fast and robust face detection. In *IEEE Transactions on pattern analysis and machine intelligence, Vol.26, NO.11*.

MacLean, W. J. and Tsotsos, J. K. (2007). Fast pattern recognition using normalized grey-scale correlation in a pyramid image representation. In *Machine Vision & Applications*.

Santiago-Mozos, R., Leiva-Murillo, J., Perez-Cruz, F., and Artes-Rodriguez, A. (1999). Supervised-pca and svm classifiers for object detection in infrared images. In *IEEE Conference on Advanced Video and Signal Based Surveillance*.

Sung, K. K. and Poggio, T. (1998). Example-based learning for view based human face detection. In *IEEE Transactions on pattern analysis and machine intelligence, Vol.22, NO.1*.

Viola, P. and Jones, M. (2001). Robust real-time object detection. In *Second International Workshop on Statistical and Computational Theories of Vision - Modeling, Learning and Sampling*.

Wakahara, T., Kimura, Y., and Tomono, A. (2001). Affine-invariant recognition of gray-scale characters using global affine transformation correlation. In *IEEE Transactions on pattern analysis and machine intelligence, Vol.23, NO.4*.

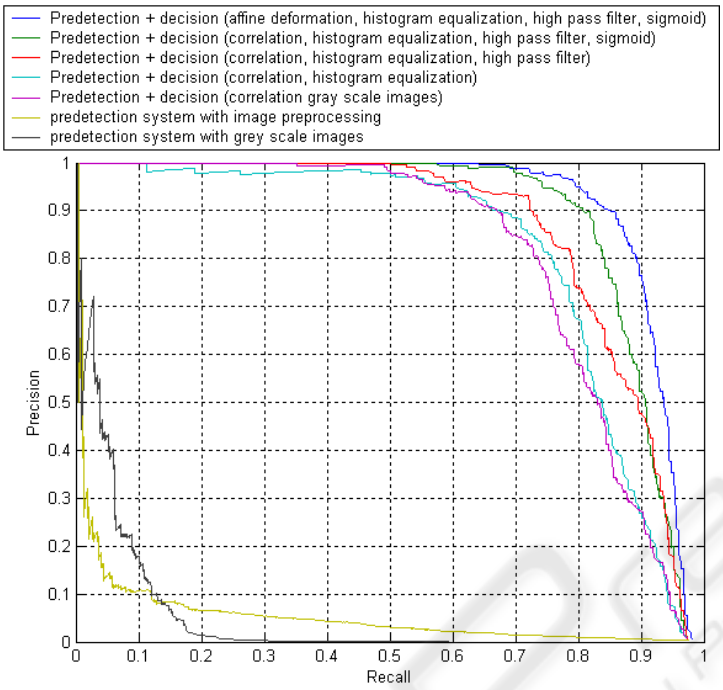


Figure 5: Relation between the precision and the recall values for differents versions of our system detection. We start from the simple predetection system, then we add the decision system using a simple grey-scale correlation, we progressively apply the different image processing to decision and finally the affine deformation method.

Figure 6: Reference images used for the system evaluation.

Figure 7: Some results obtained on the Faces 1999 database.