

LASIC: A MODEL INVARIANT FRAMEWORK FOR CORRESPONDENCE

Bernardo Rodrigues Pires, José M. F. Moura

Carnegie Mellon University
Department of ECE
Pittsburgh, PA
{bpires, moura}@ece.cmu.edu

João Xavier

Institute for Systems and Robotics
Technical University of Lisbon
Portugal
jxavier@isr.ist.utl.pt

ABSTRACT

In this paper we address two closely related problems. The first is the *object detection* problem, i.e., the automatic decision of whether a given image represents a known object or not. The second is the *correspondence* problem, i.e., the automatic matching of the features of an object in two views. We use a feature-based approach for both problems. In the first problem, we assume object rigidity and model the distortions by a linear shape model. To solve the decision problem, we derive the uniformly most powerful (UMP) hypothesis test that is invariant to the linear shape model. We use the UMP statistic to formulate the correspondence problem in a model invariant framework. We show that it is equivalent to a quadratic maximization on the space of permutation matrices. We derive LASIC, an iterative computationally feasible solution to the quadratic maximization problem for the particular case where the linear shape model is the affine model. Simulations benchmark LASIC against two standard algorithms.

Index Terms— Correspondence, Linear Shape Model, UMP-Invariant Test, Optimization methods

1. INTRODUCTION

The objective of this paper is to study two standard problems in Computer Vision: object detection and correspondence. We use a feature based approach to solve both problems. In object detection the goal is to decide if a given image corresponds to a known object. In the correspondence problem it is to establish the correspondence, or matching, between the features in two views of the same object.

Object Detection We assume: **1)** we are given a set of features that represent the shape of the known object; and **2)** if the image corresponds to the known object, the location of the features in the image can be related to the location of the known features by a parametric linear shape model. We formulate the detection problem as an hypothesis test. We require the test statistic to be invariant to the linear shape model, i.e., the result of the test should not depend on the values of the parameters that relate the features in the image to the features in the original representation. We obtain the Uniformly Most Powerful (UMP) invariant test [1] [2] and thus establish a model invariant framework. This is equivalent to the CFAR matched subspace filter developed for communication applications [2].

Correspondence Problem Correspondence is a standard problem in Computer Vision. A common assumption uses image correlation [3] [4] [5]. In not very fast videos, the solution is usually based on proximity assumptions [3]. A more general assumption is scene

rigidity [6]. Most of the methods formulate correspondence as an optimization problem and can thus be distinguished by the optimization algorithms used: greedy algorithms, linear programming via relaxation of constraints [5], randomized search [6] [4], dynamic programming [7], graph search [8] and convex optimization [3] have all been used with various degrees of success.

We apply to correspondence the model invariant framework that we develop for object detection. We model correspondence by a permutation matrix Q and show that its solution leads to an integer quadratic optimization problem. This is intractable due to its combinatorial structure. By particularizing the shape to the affine model and relaxing the search space to doubly stochastic matrices, we can design an iterative optimization algorithm for the integer quadratic problem. This is the (linear) affine shape invariant correspondence (LASIC) algorithm.

Paper Organization In Section 2, we present the solution to the object detection problem and introduce the model invariant framework. Section 3 uses this framework to formulate the correspondence problem as an integer quadratic maximization problem. Section 4 particularizes the problem to the specific structure of the affine model, and Section 5 presents LASIC, the affine optimization algorithm that solves the optimization problem. Finally, Section 6 describes experiments and Section 7 concludes the paper.

2. MODEL INVARIANT FRAMEWORK

This section focuses on the object detection problem. For clarity, we start with the affine model as an introduction to the linear shape model. We formulate the decision problem as a hypothesis test and require invariance to the linear shape model. We derive the UMP invariant test that establishes a model invariant framework.

2.1. Linear Shape Model

We assume we are given two sets of N features. The first set corresponds to a known object while the second set is extracted from a new image. The goal of the object detection problem is to determine whether the features in the new image represent an observation of the features of the known object. We follow the terminology of shape theory, e.g. [9] [10].

We use the $N \times 1$ vectors \mathbf{x}_1 and \mathbf{y}_1 to represent the x and y locations of the features in the known object. Similarly, we use the $N \times 1$ vectors \mathbf{x}_2 and \mathbf{y}_2 to represent the features in the new image. We assume object rigidity and far field so that, if the image corresponds to the known object, the location of the features in the

new image (x_2, y_2) must be related to the location of the features in the known object (x_1, y_1) by a linear shape model.

To introduce this model, we consider the affine model:

$$\begin{bmatrix} x_2^T \\ y_2^T \end{bmatrix} = \underbrace{\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}}_A \begin{bmatrix} x_1^T \\ y_1^T \end{bmatrix} + \underbrace{\begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix}}_\delta \otimes \mathbf{1} \quad (1)$$

where $\mathbf{1}$ represents a $1 \times N$ vector of ones. To obtain a linear shape model from equation (1), we re-write it as:

$$x = S\theta \quad \text{with} \quad \begin{aligned} x &= \begin{bmatrix} x_2^T & | & y_2^T \end{bmatrix}^T \\ \theta &= \begin{bmatrix} a_{11} & a_{12} & \delta_1 & a_{21} & a_{22} & \delta_2 \end{bmatrix}^T \\ S &= \left[\begin{array}{ccc|ccc} x_1 & y_1 & \mathbf{1} & & & \\ \hline & & & x_1 & y_1 & \mathbf{1} \end{array} \right] \end{aligned} \quad (2)$$

where x is the observation vector, S is the shape matrix and θ is the observation parameter vector.

Equation (2) represents the particular structure of the linear shape model in the affine case. Its general formulation is simply $x = S\theta$ where the shape matrix S is a general $n \times m$ matrix with no particular structure and x and θ are $n \times 1$ and $m \times 1$ vectors.

2.2. Object Detection: Hypothesis Test

We assume that the observation is perturbed by a vector n of i.i.d. random variables with distribution $N(\mathbf{0}, \sigma^2 I)$ and write:

$$x = \mu S\theta + n \quad (3)$$

where we use the variable $\mu \geq 0$ to define the presence ($\mu > 0$) or absence ($\mu = 0$) of the shape in the observation.

We formulate the following binary hypothesis testing problem:

Hypothesis 0: The observation is not compatible with the known shape matrix. The parameter space associated with this hypothesis is: $\mathcal{S}_0 = \{ \mu, \theta, \sigma^2 \mid \mu = 0, \theta \in \mathbb{R}^m \setminus \{\mathbf{0}\}, \sigma^2 \in \mathbb{R}^+ \}$

Hypothesis 1: The observation corresponds to the known shape matrix. The parameter space associated with this hypothesis is: $\mathcal{S}_1 = \{ \mu, \theta, \sigma^2 \mid \mu > 0, \theta \in \mathbb{R}^m \setminus \{\mathbf{0}\}, \sigma^2 \in \mathbb{R}^+ \}$

This formulation shows that the decision of whether or not the view corresponds to the given shape depends only on the parameter μ . The remaining unknowns, the observation parameters θ and the noise variance σ^2 , correspond to nuisance parameters whose presence is undesired. This motivates us to construct a decision test that is invariant to θ and σ^2 .

2.3. Model Invariant Test

We present the UMP model invariant test. This test is equivalent to the CFAR matched subspace filter [2]. We use the following model-invariant test statistic:

$$M(x) = \frac{\|P_S x\|^2/m}{\|(I - P_S)x\|^2/(n-m)} \sim \mathcal{F}\left(\frac{\mu^2 \theta^T S^T S \theta}{\sigma^2}\right) \quad (4)$$

where the notation \mathcal{F} is used to designate the \mathcal{F} distribution (with parameters m and $n - m$). Note the physical interpretation of this statistic: it measures the ratio of “energy per dimension” of the observation in the subspace $\langle S \rangle$ and in the subspace orthogonal to $\langle S \rangle$.

Using this statistic we write the UMP invariant test as:

$$\phi(M(x)) = \begin{cases} 1, & M(x) > \mathcal{F}_0 \\ 0, & M(x) \leq \mathcal{F}_0 \end{cases} \quad (5)$$

where the threshold \mathcal{F}_0 can be found by specifying the size of the test (probability of false alarm) $\alpha = P[M(x) > \mathcal{F}_0 \mid H_0 \text{ is true}]$. Under hypothesis H_0 , we have $\mu = 0 \Rightarrow M(x) \sim \mathcal{F}_{(m, n-m)}(0)$, so that we can derive \mathcal{F}_0 and design the test.

3. FEATURE CORRESPONDENCE

In section 2, we assumed that the observed image is *in correspondence* with the known shape. In computer vision problems this is seldom the case. In a typical setup, we are able to relate the observation vector x and the shape matrix S only up to a general permutation matrix Q . In this and the subsequent sections, we present our approach to effectively estimate Q .

If we knew the correct permutation matrix, Q^* , we could replace x with Q^*x in $M(x)$ in (4), which is the UMP invariant test statistic. However, the performance of the test decreases significantly if we use a $Q \neq Q^*$. Hence, we estimate the permutation matrix by finding the value of Q that maximizes the test statistic $M(Qx)$:

$$\hat{Q} = \arg \max_{Q \in \mathbb{P}} \frac{\|P_S Qx\|^2}{\|(I - P_S)Qx\|^2}, \quad (6)$$

where \mathbb{P} denotes the space of the $n \times n$ permutation matrices. Because the objective function in equation (6) is monotonically increasing with $\|P_S Qx\|^2$, we can write:

$$\hat{Q} = \arg \max_{Q \in \mathbb{P}} \|P_S Qx\|^2. \quad (7)$$

Equation (7) shows that Q can be estimated by maximizing a convex (quadratic) problem in the space of $N \times N$ permutation matrices. Due to its combinatorial nature, this problem is intractable. In the next sections, we particularize the problem to the affine shape model, for which we are able to obtain an efficient heuristic.

4. AFFINE SHAPE MODEL

In this section, we return to the affine model, and particularize the correspondence optimization problem (7) to the specific structure of this model 2. Using the structure of S and the fact that in (7) only the projector P_S matters, we can remove the means and orthonormalize the vectors x_1 and y_1 , i.e., we can make:

$$x_1 \perp y_1 \perp \mathbf{1} \quad \|x_1\| = \|y_1\| = 1 \quad (8)$$

Given the structure of x and S in (2), the permutation matrix Q is restricted to have the form:

$$Q = \left[\begin{array}{c|c} Q & \mathbf{0} \\ \hline \mathbf{0} & Q \end{array} \right] \quad (9)$$

where Q is a generic $N \times N$ permutation matrix (where N is the number of features). With the structure in (9), and after simple manipulations, we can write the objective function in (7) (the “energy” of x in the shape plane) as:

$$f(q) = \|P_S Qx\|^2 = \sum_{i=1}^4 \left(u_i^T q \right)^2 \quad (10)$$

where we introduced the $N^2 \times 1$ vectors $q = \text{vec}(Q)$, and $u_1 = x_1 \otimes x_2$, $u_2 = y_1 \otimes x_2$, $u_3 = x_1 \otimes y_2$, $u_4 = y_1 \otimes y_2$.

Note that the objective function (10) measures the sum of the “energy” of the projection of the vector q into four known vectors.

5. LASIC: MAXIMIZATION ALGORITHM

In Section 3, we formulated the correspondence problem as an integer quadratic maximization problem. In section 4, we showed that for the affine shape model, a special structure appeared in the objective function (10). In this section, we explore this structure to derive the (linear) affine shape invariant correspondence (LASIC) algorithm, an efficient iterative heuristic for the problem.

5.1. Relaxation: Doubly Stochastic Matrices

Our goal is to maximize the objective function in (10) with respect to $\mathbf{q} = \text{vec}(\mathbf{Q})$ where $\mathbf{Q} \in \mathbb{P}$, the space of the $N \times N$ permutation matrices. We introduce the set of the $N \times N$ doubly stochastic matrices, \mathbb{DS} , defined as:

$$0 \leq \mathbf{Q}_{ij} \quad , \quad \sum_i \mathbf{Q}_{ij} \leq 1 \quad , \quad \sum_j \mathbf{Q}_{ij} \leq 1. \quad (11)$$

Birkhoff's theorem [11] states that the set of $N \times N$ doubly stochastic matrices is a compact convex set whose extreme points are the $N \times N$ permutation matrices. Because we are maximizing a convex (quadratic) function over a compact convex set, \mathbb{DS} , the solution will be an extreme point of the set and so in \mathbb{P} . Thus, relaxing the problem from the set \mathbb{P} to \mathbb{DS} does not change its solution, i.e.,

$$\max_{\substack{\mathbf{q}=\text{vec}(\mathbf{Q}) \\ \mathbf{Q} \in \mathbb{P}}} \sum_{i=1}^4 (\mathbf{u}_i^T \mathbf{q})^2 \Leftrightarrow \max_{\substack{\mathbf{q}=\text{vec}(\mathbf{Q}) \\ \mathbf{Q} \in \mathbb{DS}}} \sum_{i=1}^4 (\mathbf{u}_i^T \mathbf{q})^2 \quad (12)$$

The proposed LASIC algorithm solves this in two steps: an initialization algorithm and an iterative optimization procedure. The initialization algorithm generates an initial solution that is then used by the iterative optimization procedure. After multiple re-initializations, the limiting point that achieves the largest value of the objective function is taken as the maximum.

5.2. LASIC Initialization Procedure

To obtain an initial vector \mathbf{q}_0 , we approximate problem (12) by:

$$\max_{\substack{\mathbf{q}=\text{vec}(\mathbf{Q}) \\ \mathbf{Q} \in \mathbb{DS}}} \sum_{i=1}^4 |\mathbf{u}_i^T \mathbf{q}|, \quad (13)$$

This is a sum of four absolute values. Each one can be replaced with its original value or its symmetric. We solve for all the 16 (2^4) combinations of positive and negative factors, thus writing 16 linear optimization problems of the form:

$$\max \sum_{i=1}^4 (-1)^{k_i} \mathbf{u}_i^T \mathbf{q} \quad , \quad k_i = \{0, 1\}, \quad (14)$$

subject to $\mathbf{q} = \text{vec}(\mathbf{Q})$, $\mathbf{Q} \in \mathbb{DS}$ and $(-1)^{k_i} \mathbf{u}_i^T \mathbf{q} \geq 0$ with $i = 1, 2, 3, 4$. Writing the objective function as a sum of only four factors is a key step to obtain an efficient method, since the number of linear programs to solve grows exponentially with the number of factors in (13). After solving all the 16 problems, the solution that attains the maximum of the objective function is taken as \mathbf{q}_0 .

5.3. LASIC Iterative Optimization Procedure

After the initialization, in subsection 5.2, we use an iterative algorithm to solve problem (10). Since (10) is quadratic, convex upwards, given an estimate \mathbf{q}_k of \mathbf{q} , we can use the tangent to lower bound $f(\mathbf{q})$:

$$f(\mathbf{q}) \geq f(\mathbf{q}_k) + \nabla_{\mathbf{q}}^T f(\mathbf{q}_k) \cdot (\mathbf{q} - \mathbf{q}_k), \quad (15)$$

where ∇ is the gradient operator. To maximize $f(\mathbf{q})$, we maximize this lower bound. This leads to the LASIC iterative algorithm:

$$\mathbf{q}_{k+1} = \arg \max_{\substack{\mathbf{q}=\text{vec}(\mathbf{Q}) \\ \mathbf{Q} \in \mathbb{DS}}} \left(\sum_{i=1}^4 (\mathbf{u}_i^T \mathbf{q}_k) \mathbf{u}_i^T \right) \mathbf{q}. \quad (16)$$

The process is repeated until a convergence criterium is met.

5.4. Multiple Initializations

The LASIC iterative maximization algorithm (16) is not guaranteed to converge to the absolute maximum of the objective function. We run it with several initializations and keep the overall maximum. The question is how to generate these several initializations. We observe that the original problem is (12) is invariant to rotations in the vectors $(\mathbf{x}_1, \mathbf{y}_1)$ but the LASIC initialization algorithm (13) is not. So we re-initialize the algorithm with a new set of vectors $(\mathbf{x}_1, \mathbf{y}_1)$ that is a rotation of the original set $(\mathbf{x}_1, \mathbf{y}_1)$.

6. EXPERIMENTAL RESULTS

We benchmark LASIC against the proposed method to two correspondence algorithms: image correlation correspondence [5] and feature location correspondence. These algorithms attempt to solve the problem by maximizing measures of correlation between the features. The first method extracts an image patch (we use a 21×21 square) around each feature and maximizes the correlation in the image values. The second method maximizes the correlation in the feature locations using an objective function similar to [12]. Like LASIC, both comparison algorithms formulate correspondence as an optimization problem in the space of permutation matrices.

We develop three synthetic situations to compare the performance of the algorithms: addition of Gaussian noise to the location of the features, isotropic zoom in the object, and rotation of the object. We measure the performance of the algorithms by the average of the correspondence error rate (CER) defined as the ratio of mismatches to the total number of features. In all experiences, the number of features is set to 30 and LASIC is re-initialized 20 times.

Figures 2, 3, and 4 summarize the results. Figure 1 shows the features of the known object. Figure 2 plots CER versus the noise standard deviation, σ . Figure 3 displays CER versus the isotropic zoom parameter, z , for a constant noise level. Note that the isotropic zoom can be modeled by the affine model (1) where \mathbf{A} is a scalar times the identity and there is no translation. Finally, figure 4 illustrates the CER versus the rotation parameter, θ . Again, the rotation can be modeled by an affine model (1) where \mathbf{A} is a rotation matrix and there is not translation.

The image correlation correspondence is known to be very robust to noise added on the image pixel values [5], while here the noise is added to the feature location. Our experiments show that it fails when noise is added directly to the feature locations, or when either zoom or image rotation occur.

Figures 2 and 3 show that LASIC is robust under noise and zoom, its performance is slightly worse than the location correspondence method. Further exploration into the zoom experience is needed in order to determine if we are preventing full convergence by restricting the number of initializations to 20. Rotation demonstrates the strength of LASIC: while the comparison algorithms, especially the location correspondence algorithm, break down even for small rotations, LASIC obtains nearly perfect results.

7. CONCLUSION

We considered the problem of deciding if a given image corresponds to a known object. We took a feature based approach and modeled the distortions between the new image and the known object by a linear shape model. We presented an invariant framework and developed the corresponding UMP test. We then formulated the correspondence problem in this invariant framework as a quadratic maximization on the space of permutation matrices. For affine shapes, we can derive an efficient iterative optimization algorithm, the (linear) affine shape invariant optimization (LASIC) algorithm. Experiments showed that LASIC handles situations where other methods fail, e.g., large rotations.

8. REFERENCES

- [1] E.L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses*, Springer, third edition, 2005.
- [2] L.L. Scharf, *Statistical Signal Processing: Detection, Estimation and Time Series Analysis*, Addison-Wesley, 1991.
- [3] B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. of the 7th International Joint Conference on Artificial Intelligence*, 1981, pp. 674–679.
- [4] P. Torr and D. Murray, "The development and comparison of robust methods for estimating the fundamental matrix," *Intl. Journal of Computer Vision*, vol. 24, no. 3, pp. 271–300, 1997.
- [5] J. Maciel and J. P. Costeira, "A global solution to sparse correspondence problems," *Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 2, pp. 139154, 2003.
- [6] G. Sudhir, S. Banerjee, and A. Zisserman, "Finding point correspondences in motion sequences preserving affine structure," *Computer Vision and Image Understanding*, vol. 68, no. 2, pp. 237–246, 1997.
- [7] Y. Ohta and T. Kanade, "Stereo by intra- and inter-scanline search using dynamic programming," *Tr. on Pattern Analysis and Machine Intelligence*, vol. 7, no. 2, pp. 139154, 1985.
- [8] S. Roy and I. Cox, "A maximum-flow formulation of the n-camera stereo correspondence problem," in *Proc. of the International Conference on Computer Vision*, 1997, pp. 492–502.
- [9] T. K. Carne D. G. Kendall, D. Barden and H. Le, *Shape and Shape Theory*, Wiley, first edition, 1999.
- [10] Victor Ha and Jos M. F. Moura, "Affine-permutation invariance of 2d shapes," *Transactions on Image Processing*, vol. 14, no. 11, pp. 1687–1700, 2005.
- [11] D.R. Horn and C. Johnson, *Matrix Analysis*, Cambridge U. Press, 1985.
- [12] E.Z. Psarakis and G.D. Evangelidis, "An enhanced correlation-based method for stereo correspondence with subpixel accuracy," in *Proc. of the International Conference on Computer Vision*. IEEE, October 2005, pp. 907 – 912.

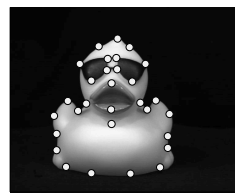


Fig. 1. Original image

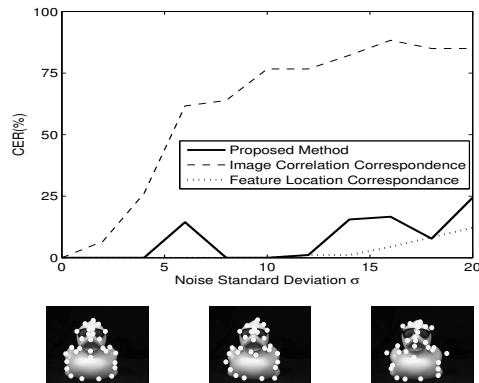


Fig. 2. Adding noise to feature locations

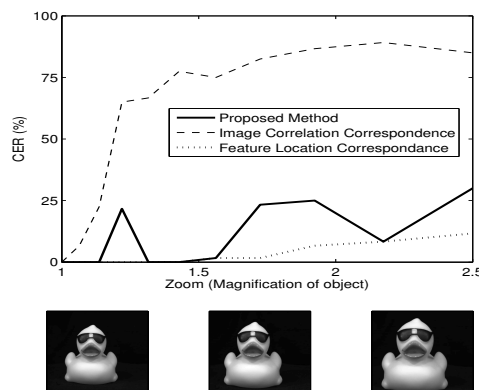


Fig. 3. Zooming in on the object

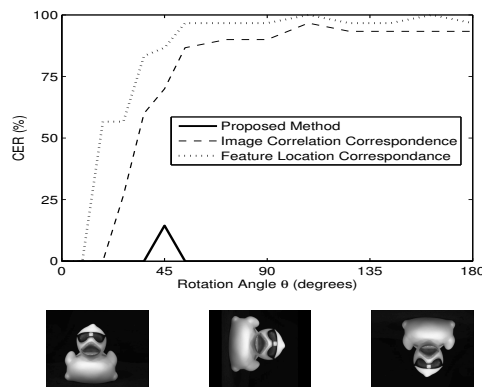


Fig. 4. Rotating the object