GROUP-WISE SPARSE CORRESPONDENCES BETWEEN IMAGES BASED ON A COMMON LABELLING APPROACH

Albert Solé-Ribalta[1], Gerard Sanromà[1], Francesc Serratosa and René Alquézar[2]
[1]*Departament d'Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili, Tarragona, Spain*
[2]*Institut de Robtica i Informàtica Industrial, CSIC-UPC, Barcelona, Spain*
{albert.sole, gerard.sanroma, francesc.serratosa}@urv.cat, ralquezar@iri.upc.edu

Keywords: Multiple Point Set Alignment, Group Wise Point Set Alignment.

Abstract: Finding sparse correspondences between two images is a usual process needed for several higher-level computer vision tasks. For instance, in robot positioning, it is frequent to make use of images that the robot captures from their cameras to guide the localisation or reduce the intrinsic ambiguity of a specific localisation obtained by other methods. Nevertheless, obtaining good correspondence between two images with a high degree of dissimilarity is a complex task that may lead to important positioning errors. With the aim of increasing the accuracy with respect to the pair-wise image matching approaches, we present a new method to compute group-wise correspondences among a set of images. Thus, pair-wise errors are compensated and better correspondences between images are obtained. These correspondences can be used as a less-noisy input for the localisation process. Group-wise correspondences are computed by finding the common labelling of a set of salient points obtained from the images. Results show a clear increase in effectiveness with respect to methods that use only two images.

1 INTRODUCTION

Determining sparse correspondences between sets of features is a recurrent problem in computer vision. It arises at the early stages of many computer vision applications such as 3D scene reconstruction, object recognition, pose recovery and image retrieval, among others. Therefore, it is of basic importance to develop methods that are both effective -in the sense of not being prone to local optima- and robust -in the sense of being able to accommodate a wide range of image deformations as well as noisy measurements-. We divide classical approaches to compute pair-wise correspondences into: (1) correlation-based strategies that compute the matches by means of the similarity between the image patches around some interest points (Harris and Stephens 1988) and; (2) approaches based on feature-descriptors that use local information at the interest points to compute descriptor-vectors (Mikolajczyk and Schmid 2005).

The use of local image contents may not suffice to get a reliable correspondence between points of two images under certain circumstances e.g. large rigid/non-rigid deformations. This is the case of the model fitting paradigm RANSAC (Fischler and Bolles 1981) which is extensively used in computer vision to reject outliers or the Iterative Closest Point (ICP) method (ZHANG 1992) that attempt to simultaneously solve the correspondence and the alignment problem. All the mentioned approaches suffer from two major drawbacks. On the one hand, most of these optimization strategies rely on reasonable initial guesses in order to find the global optimum. On the other hand, if there is too much deformation between both images, their underlying geometrical models may fail to accommodate the transformation relating them, even under a reasonable initial guess.

To solve the aforementioned drawbacks, we face the correspondence problem in a group-wise manner. In this way, the flow of information among the pair-wise relations of the group has several advantages. It helps to constrain the search of our method towards a globally convenient direction. This contributes to avoid poor local optima. And, in addition it alleviates the limitations inherent to the geometrical models. To complement the method, we develop effective mechanisms to detect outlying points between two point-sets whose effects are conveniently propagated to the rest of the group.

The approach we propose has been successfully applied to graph matching (Solé-Ribalta

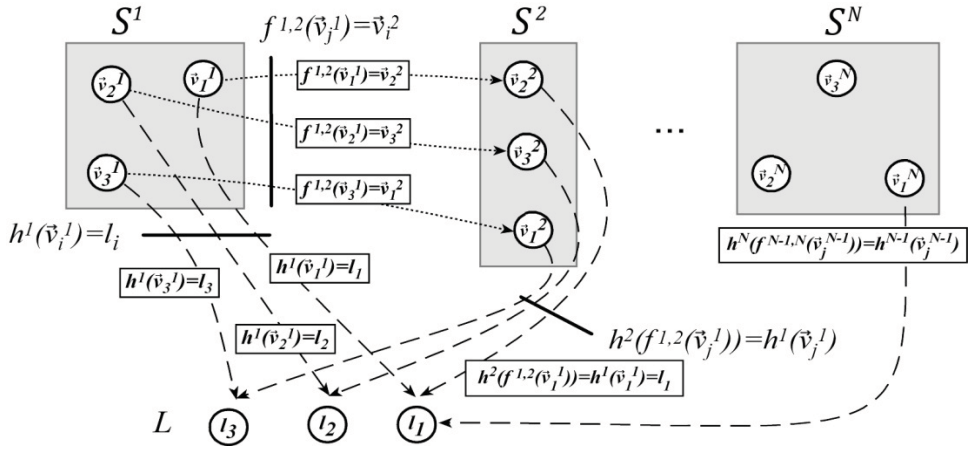

Figure 1: Example of common labelling.

and Serratosa 2010). Here, we adapt the graph-oriented solution to group-wise point registration and we enhance its effectiveness incorporating a geometrical model and an outlier detection mechanism.

Several similar solutions have been proposed for graph matching purposes. We highlight (Williams, Wilson et al. 1997) where some pair-wise matchings where induced using Bayesian inference. The main limitation of the methodology is that is not applicable to more than 3 graphs. Another solution, also applied to graphs, was proposed in (Solé-Ribalta and Serratosa 2011). In this case the extension to multiple graphs is straightforward; however, its high computational cost makes it again not applicable with more than 3 graphs.

Related to the field of group-wise point registration when data is a sparse set of points we next highlight the following work. In (Fergus, Perona et al. 2007) a method to learn objects and detect parts of objects is presented. The model is learned taking images that represent the selected object from the same point of view and without background. The method does not explicitly address the problem presented here due to the aim is to construct a model for object recognition. Another related work is presented in (Wang, Vemuri et al. 2008), which performs alignment of sparse data points taking into account that points contain non-rigid deformation. The most similar method to the one we present could be (Cootes, Twining et al. 2010). It is based on group-wise point set correspondence but it has no consideration about outlier detection, which makes its applicability not feasible with the concrete problem we present. This last work was evaluated using two hand-made labelled data sets.

The article is structured as follows. Section 2 gives some basic definitions. Section 3 and 4 describes the method used to deduct the group-wise correspondences. Section 3 describes the common labelling framework and section 4 describes the cost function as well as the outlier detection procedure. Section 5 describes the optimization algorithm. Section 6 evaluates the new method and Section 7 concludes the paper.

2 BASIC DEFINITIONS

Let $S^p = \{\vec{v}_1^p, \vec{v}_2^p, .. \vec{v}_{O^p}^p\}$ be a set of points with O^p elements. In our method, these types of sets represent images and their elements are salient points extracted from them. Moreover, we represent the set of images by the set $\Gamma=\{S^1, S^2, ..., S^N\}$. Each S^p in Γ is the characterisation of an image. Following this notation, the correspondence between salient points of a set of images are characterised by the labellings between the elements of the sets S^p in Γ. Note that outlier points in images are also represented as elements in S^p. These outlier points in the images do not correspond to other points on the other images and so the corresponding elements in the sets have not to be labelled from or to these elements.

Definition 1. Labelling between two sets of points: Given two sets of points $S^p = \{\vec{v}_1^p, \vec{v}_2^p, .., \vec{v}_{O^p}^p\}$ and $S^q = \{\vec{v}_1^q, \vec{v}_2^q, .., \vec{v}_{O^q}^q\}$ with O^p and O^q elements, a labelling f between these sets assign elements of the first set to elements of the second set $\left(\vec{v}_a^p\right) = \vec{v}_i^q$. We represent this labelling in a binary matrix as follows,

$$M_{ai} = \begin{cases} 1 \; if \; f(\vec{v}_a^p) = \vec{v}_i^q \\ 0 \; otherwise \end{cases} \quad (1)$$

Definition 2. Multiple Labelling between sets of points. Let $\Gamma=\{S^1, S^2, ..., S^N\}$ be a set of N sets of points, each with a concrete number of elements O^p, $p: 1..N$. The set φ is a Multiple Labelling of Γ if it contains one and only one labelling between any pair of set of points, $\varphi = \{f^{1,2}, \dots, f^{2,1}, \dots, f^{N-1,N}\}$.

Some inconsistencies may appear in the multiple labelling φ if these labellings are obtained by only considering the set of points they relate. Fig. 2.a and Fig. 2.b shows an example of an inconsistent and

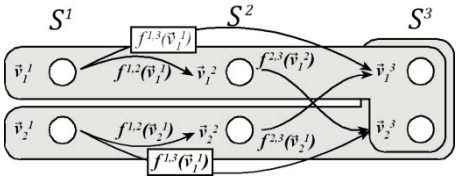

Figure 2.a: Example of inconsistent multiple labelling.

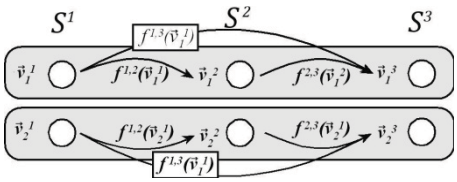

Figure 2.b: Example of consistent multiple labelling.

consistent multiple labelling respectively. In the inconsistent labelling point $\vec{v}_1^1$ is labelled to $\vec{v}_1^2$ by function $f^{1,2}$ and it is labelled to $\vec{v}_1^3$ by $f^{1,3}$. However, $f^{2,3}$ labels $\vec{v}_1^1$ to $\vec{v}_2^3$. Therefore, there is no a global correspondence between the salient points in the original images. See how this is fixed in the consistent multiple labelling.

Some methods solve this problem by first finding the pair-wise labelling to next redefining them with the aim of eliminating inconsistencies (Bonev, Escolano et al. 2007). The main property of our method is that it obtains directly a multiple labelling. That is, it considers from the first moment the group-wise correspondences, and so, inconsistencies cannot appear due to our methodology.

We say that a multiple labelling is consistent if there are not any inconsistencies. That is, it fulfils that,

$$f^{q,k}\left(f^{p,q}(\vec{v}_i^p)\right) = f^{p,k}(\vec{v}_i^p), \quad 0 < p,q,k \le N, 0 < i \le O^p \quad (2)$$

When the sets of points in Γ form a Consistent Multiple Labelling, it is possible to define a Common Labelling. It represents a Multiple Labelling in a compact way and forms the basis of the proposed algorithm.

Definition 3. Common Labelling (CL) of sets of points. Let φ be a Consistent Multiple Labelling of Γ. Let L be a virtual point set. The Common Labelling $\psi = \{h^1, h^2, \dots, h^N\}$ is defined to be a set of bijective mappings from the points of Γ to L as follows:

$$h^1(\vec{v}_i^1) = i, h^p(\vec{v}_i^p) = h^{p-1}(\vec{v}_j^{p-1}), \quad 1 \le i,j \le O^p, 2 \le p \le N, \quad being \; f^{p-1,p}(\vec{v}_j^{p-1}) = \vec{v}_i^p \quad (3)$$

Figure 1 shows the relation between a Consistent Multiple Labelling and a Common Labelling.

3 COMMON LABELLING FRAMEWORK

Given two sets of salient points, S^p and S^q, extracted from two images, to bring the problem to the continuous domain we relax the matches between these point sets i.e. (1). To this aim, we represent the probability of labelling S^p to S^q in matrix from as:

$$P_f^{p,q}[i,j] = Prob(f^{p,q}(\vec{v}_i^p) = \vec{v}_j^q) \quad (4)$$

Moreover, we consider the probability of labelling point $\vec{v}_i^p$ of set S^p to a virtual point l_j is the probabilistic union of all the paths that go through the points of a third set S^q. That is,

$$P_h^p[i,j] = Prob(h^p(\vec{v}_i^p) = l_j) = = Prob\left(\bigcup_{k=1}^{O^q}[f^{p,q}(\vec{v}_i^p) = v_k^q \cap h^q(\vec{v}_k^q) = l_j]\right) \quad (5)$$

Combining (5) with P_f definitions and assuming independence of events we get:

$$P_h^p[i,j] = \sum_{k=1}^{O^q} P_f^{p,q}[i,k] \cdot P_h^q[k,j] \quad or \; P_h^p = P_f^{p,q} \cdot P_h^q \quad (6)$$

In a similar way, we could infer that:

$$P_f^{p,q} = P_h^p \cdot (P_h^q)^T \quad (7)$$

Due to our final objective is to compute a CL, our new energy function depends on the probabilities P_h

instead of P_f. To this aim, we define the energy of a group-wise point alignment as:

$$E^{CL} = -\sum_{p=1}^{N}\sum_{\substack{q=1\\q\neq p}}^{N}\sum_{a=1}^{O^p}\sum_{i=1}^{O^q}\sum_{b=1}^{O^p}\sum_{j=1}^{O^q}\left(\underbrace{\sum_{w_1=1}^{O^L} P_h^p[a,w_1]\cdot P_h^q[i,w_1]}_{\equiv P_f^{p,q}[a,i]} \cdot \underbrace{\left[\sum_{w_2=1}^{O^L} P_h^p[b,w_2]\cdot P_h^q[j,w_2]\right]}_{\equiv P_f^{p,q}[b,j]} \cdot C_{ai,bj}^{p,q}\right) \quad (8)$$

Energy of (8) is a generalization of energy of pair-wise labellings.

Reorganizing (8) we can easily see the influence of matchings $\vec{v}_b^p \rightarrow \vec{v}_j^q$ over $\vec{v}_a^p \rightarrow \vec{v}_i^q$:

$$E^{CL} = -\sum_{p=1}^{N}\sum_{\substack{q=1\\q\neq p}}^{N}\sum_{a=1}^{O^p}\sum_{i=1}^{O^q}\left(\underbrace{\sum_{w_1=1}^{O^L} P_h^p[a,w_1]\cdot P_h^q[i,w_1]}_{\equiv P_f^{p,q}[a,i]}\right) \left(\underbrace{\sum_{b=1}^{O^p}\sum_{j=1}^{O^q}\underbrace{\left[\sum_{w_2=1}^{O^L} P_h^p[b,w_2]\cdot P_h^q[j,w_2]\right]}_{\equiv P_f^{p,q}[b,j]}\cdot C_{ai,bj}^{p,q}}_{R_{ai}^{pq}}\right) \quad (9)$$

This influence is identified as R_{ai}^{pq} in (9) and will be described in detail in section 4.

4 PAIR-WISE COMPATIBILITY COEFFICIENTS

Given two sets of points $S^p = \{\vec{v}_1^p, \vec{v}_2^p, .., \vec{v}_{O^p}^p\}$ and $S^q = \{\vec{v}_1^q, \vec{v}_2^q, .., \vec{v}_{O^q}^q\}$, where $\vec{v}_k^p = \left(v_{k_H}^p, v_{k_V}^p\right)^T$ and $\vec{v}_l^q = \left(v_{l_H}^q, v_{l_V}^q\right)^T$, contain the column vectors of the two-dimensional coordinates (horizontal and vertical) of each point, in this section we will describe the details of the computation of the compatibility coefficients $C_{ai,bj}$ appearing under equation (9).

This quantity R_{ai}^{pq}, also known as the support function, is addressed at measuring the support for the match $\vec{v}_a^p \rightarrow \vec{v}_i^q$ received from the rest of the matches $\vec{v}_b^p \rightarrow \vec{v}_j^q$. This is a common strategy followed in the probabilistic relaxation approaches (Rosenfeld, Hummel et al. 1976; Hummel and Zucker 1983).

The main idea underpinning our computation of the support function is that two points $\vec{v}_a^p$ and $\vec{v}_i^q$ from two graphs p and q are in correspondence as long as they show similar spatial distributions in comparison to the rest of the points around them.

Geometric evidence is widely used to solve the correspondence problem. In order to be robust to arbitrary initial poses of the point-sets under a certain geometric assumption we need to include the estimation of the alignment parameters into the problem. Thus we redefine the support function in the following way

$$R_{ai}^{pq} = \max_{\Phi_{ai}} \sum_{b=1}^{O^p}\sum_{j=1}^{O^q} P_f^{p,q}[b,j]\cdot C_{ai,bj}(\Phi_{ai}) \quad (10)$$

where $P_f^{p,q}[b,j]$ corresponds to the globally propagated probability to match nodes b, j of graphs p, q and $C_{ai,bj}(\Phi_{ai})$ is the compatibility of the simultaneous matches $\vec{v}_a^p \rightarrow \vec{v}_i^q$ and $\vec{v}_b^p \rightarrow \vec{v}_j^q$ given the affine parameters Φ_{ai}.

In this new formulation, we attain robustness to affine pose of the point-sets by selecting the pose configuration that leads to the maximum support.

With respect to classical point-set registration methods, our approach has the particularities that it is aimed at multiple point-set registration and that alignment parameters are local to each correspondence hypothesis $v_a^p \longrightarrow v_i^q$ instead of being a property global to all the points in the set.

Since we compare relational geometric measurements, we define the new coordinate vectors $\vec{x}_{ab} = \left(\vec{v}_b^p - \vec{v}_a^p\right)$ and $\vec{y}_{ij} = \left(\vec{v}_j^q - \vec{v}_i^q\right)$, that represent the coordinates of the points $\vec{v}_b^p$ and $\vec{v}_j^q$ relative to $\vec{v}_a^p$ and $\vec{v}_i^q$, respectively.

We define the compatibility between two relational geometric measurements $\vec{x}_{ab}$ and $\vec{y}_{ij}$ under the action of the affine parameters Φ_{ai} as:

$$C_{ai,bj}(\Phi_{ai}) = \rho - \left\|\vec{x}_{ab} - \Phi_{ai}\vec{y}_{ij}\right\|_{\Sigma}^2 \quad (11)$$

where Φ_{ai} is a $2x2$ non-singular matrix of affine transformation parameters (note that $\vec{x}_{ab}$ and $\vec{y}_{ij}$ are already invariant to translation), $\|\cdot\|_{\Sigma}^2$ is the squared Mahalanobis distance with covariance matrix Σ, and ρ is a thresholding quantity that controls the outlier process whose estimation will be detailed in the next section.

According to the proposed measure, the more dissimilar are the relations, the lower is their compatibility. The scale of this comparison is

effectively controlled by the matrix $\Sigma = \begin{bmatrix}\sigma_H^2 & 0\\ 0 & \sigma_V^2\end{bmatrix}$, a diagonal matrix of variances which may be empirically estimated from the data.

With these ingredients, the optimal transformation parameters Φ_{ai}^* that maximize equation (10) are:

$$\Phi_{ai}^* = \min_{\widehat{\Phi}_{ai}} \sum_{b=1}^{O^p}\sum_{j=1}^{O^q} P_f^{p,q}[b,j]\left(\vec{x}_{ab} - \widehat{\Phi}_{ai}\vec{y}_{ij}\right)^T \Sigma^{-1}\left(\vec{x}_{ab} - \widehat{\Phi}_{ai}\vec{y}_{ij}\right) \quad (12)$$

where $\Phi_{ai}^* = \begin{bmatrix}\phi_{11} & \phi_{12}\\ \phi_{21} & \phi_{22}\end{bmatrix}$. We have discarded the constant quantities not depending on the alignment parameters. Note that we have turned the maximization into a minimization by reversing the sign.

Consider the following residuals from the alignment of points $\vec{x}_{ab}$ and $\vec{y}_{ij}$.

$$\begin{aligned} r_{bj}^V &= x_{ab}^V - \left(\phi_{11}y_{ij}^V + \phi_{12}y_{ij}^H\right)\\ r_{bj}^H &= x_{ab}^H - \left(\phi_{21}y_{ij}^V + \phi_{22}y_{ij}^H\right)\end{aligned} \quad (13)$$

Then, the objective function of equation (12) is equivalent to the following expression:

$$\mathcal{F} = \sum_{b=1}^{O^p}\sum_{j=1}^{O^q} P_f^{p,q}[b,j]\left[\left(\frac{r_{bj}^V}{\sigma_H}\right)^2 + \left(\frac{r_{bj}^H}{\sigma_V}\right)^2\right] \quad (14)$$

Taking derivatives of $\mathcal{F}$ with respect to Φ_{ai} we obtain the following expressions:

$$\begin{aligned} \frac{\delta\mathcal{F}}{\delta\phi_{11}} &= -\sum_{b=1}^{O^p}\sum_{j=1}^{O^q} P_f^{p,q}[b,j]\frac{y_{ij}^V 2r_{bj}^V}{\sigma_H^2}\\ \frac{\delta\mathcal{F}}{\delta\phi_{12}} &= -\sum_{b=1}^{O^p}\sum_{j=1}^{O^q} P_f^{p,q}[b,j]\frac{y_{ij}^H 2r_{bj}^V}{\sigma_H^2}\\ \frac{\delta\mathcal{F}}{\delta\phi_{21}} &= -\sum_{b=1}^{O^p}\sum_{j=1}^{O^q} P_f^{p,q}[b,j]\frac{y_{ij}^V 2r_{bj}^H}{\sigma_H^2}\\ \frac{\delta\mathcal{F}}{\delta\phi_{22}} &= -\sum_{b=1}^{O^p}\sum_{j=1}^{O^q} P_f^{p,q}[b,j]\frac{y_{ij}^H 2r_{bj}^H}{\sigma_H^2}\end{aligned} \quad (15)$$

The optimal transformation parameters Φ_{ai}^* are found by solving the set of equations:

$$\frac{\delta\mathcal{F}}{\delta\phi_{11}} = 0, \ldots, \frac{\delta\mathcal{F}}{\delta_{22}} = 0, \quad (16)$$

with respect to the parameters. This linear system can be expressed in matrix form $M\vec{a} = \vec{b}$, where M is a $4x4$ matrix and $\vec{a} = \left(\phi_{11}, \phi_{12}, \phi_{21}, \phi_{22}\right)^T$ and $\vec{b}$ are 4-column-vectors. This can be solved by matrix inversion (i.e., $\vec{a} = M^{-1}\vec{b}$).

4.1 Outlier Detection

According to our purposes, a point $\vec{v}_a^q \in S^p$ (or $\vec{v}_i^q \in S^q$) is considered an outlier as far as there is no point $\vec{v}_i^q, \forall\, i \in 1..|S^q|$ (or $\vec{v}_a^p, \forall\, a \in 1..|S^p|$) which presents a support R_{ai}^{pq} (10) above a given threshold.

Substituting the compatibilities of equation (11) into equation (10), the final expression for the supports is:

$$R_{ai}^{pq} = \sum_{b=1}^{O^p}\sum_{j=1}^{O^q} P_f^{p,q}[b,j]\left(\rho - \left\|\vec{x}_{ai} - \Phi_{ai}^*\vec{y}_{ij}\right\|_\Sigma^2\right) \quad (17)$$

where Φ_{ai}^* are the optimal transformation parameters computed using equation (12).

The parameter ρ plays the role of the *robustness parameter* used by (Rangarajan, Chui et al. 1997; Gold, Rangarajan et al. 1998). It controls whether the geometrical compatibility term contributes either positively (i.e., $\rho < \left\|\vec{x}_{ai} - \Phi_{ai}^*\vec{y}_{ij}\right\|_\Sigma^2$) or negatively to the support measure.

We model the outlier detection process as an assignment to (or from) a special point. This is similar to *null* vertex assignments in (Wong and You 1985). We consider as outliers all the assignments $v_a^p \longrightarrow v_i^q$ such that $R_{ai}^{p,q} < 0$.

The threshold ρ represents the quantity from which the compatibility starts to contribute negatively. Therefore, it seems reasonable to express ρ in terms of a squared Mahalanobis distance, i.e. $\rho = \left\|\vec{d}\right\|_\Sigma^2$. If we express the threshold distance vector proportionally to the standard deviations of the data, i.e. $\vec{d} = (N\sigma_H, N\sigma_v)$, the expression of ρ becomes

$$\rho = \vec{d}^T\Sigma^{-1}\vec{d} = \left(\frac{N\sigma_H}{\sigma_H}\right)^2 + \left(\frac{N\sigma_H}{\sigma_H}\right)^2 = 2N^2 \quad (18)$$

considering that Σ matrix is diagonal.

Rangarajan *et al. (Rangarajan, Chui et al. 1997; Gold, Rangarajan et al. 1998)* do not address the estimation of this parameter in their paper. On the contrary, we define ρ as a function of the number N of standard deviations permitted in the registration errors, in order to consider a relation plausible.

5 THE ALGORITHM

Considering the optimization function in (9) for multiple point set matching, we focus on substituting to it the support function deduced in section 4.

The problem becomes then one of joint estimation of correspondence and alignment parameters in which the recovery of the correspondences is influenced by the pose of the point-sets and vice-versa. Most point-set registration methods consist of an iterative process that alternates alignment and correspondence updates. Several approaches exist in order to solve this chicken-and-egg problem as, for example, the well-known ICP (ZHANG 1992), Robust Point Matching (RPM) (Rangarajan, Chui et al. 1997; Gold, Rangarajan et al. 1998) or the Expectation-Maximization Algorithm (Jian and Vemuri 2005; Myronenko and Song 2010; Horaud, Forbes et al. 2011; Jian and Vemuri 2011).

To optimize our objective function we propose to use a similar dual step solution based on first maximize the point alignment to later maximize the correspondences. We base our method on the Graduated assignment (Gold and Rangarajan 1996). In this way, we approximate E^{CL} with Taylor series expansion considering that the point alignment given by Φ^*_{ai} is already optimized. Similarly to (Gold and Rangarajan 1996) we deduce that minimizing function (9) is equivalent to maximizing:

$$argmin\{E^{CL}\} = argmax\left\{\sum_{a=1}^{O^p}\sum_{w_1=1}^{O^L} Q^p_{a,w_1}\cdot P^p_h[a,w_1]\right\} \quad (19)$$

where,

$$Q^p_{a,w_1} = \sum_{q=1}^{N}\sum_{i=1}^{O^q} P^q_h[i,w_1]\left(\underbrace{\sum_{\substack{b=1\\ b\neq a}}^{O^p}\sum_{\substack{j=1\\ i\neq j}}^{O^q}\left[\sum_{w_2}^{O^L}\left(P^p_h[b,w_2]\right)^t\cdot P^q_h[j,w_2]\right]\cdot C^{p,q}_{ai,bj}}_{\kappa^{pq}_{ai}}\right) \quad (20)$$

Equation (20) reduces the problem to the quadratic assignment problem, where Q is the cost matrix and P represents a stochastic matrix (Sinkhorn 1964) that encode the assignment probabilities.

The original procedure to optimize equation (20) is the following: start with a valid $(P^p_f)^t$, compute cost matrix Q^p, apply Graduated Assignment to compute next $(P^p_f)^{t+1}$ and start again until convergence is reached.

Similarly, in our objective function to maximize the common labelling assignments we focussed in P^p_h as (8) and (9) indicates. In addition, we are required to maximize the alignment to compute the compatibility cost. So, our proposed maximization procedure has the following steps: start with a valid $(P^p_h)^t$, maximize alignment with respect to the rest of points (12), compute cost matrix Q^p using costs in (10), apply Graduated Assignment to compute next $(P^p_h)^{t+1}$ and start again until convergence is reached. An outline of the procedure is given below.

Program MSP-Aligment **inputs** Γ **returns** P^p_h
  $Initialize\ P_h$
  $\beta = \beta_0$
  $\rho = 2N^2$
  Loop A: (**Do** A **until** $\beta \geq \beta_f$)
    $it = 0$
    Loop B: (**Do** B **until** Q converges **or** $it < I_0$)
      $Q^p_{a,W_1} = ComputeQ(P_h, p, a, w_1, \Gamma)$
      $P^p_h = e^{\beta\cdot Q^p}$
      $P^p_h = sinkhorn(P^p_h)$
      $it = it + 1$
    End B
    $\beta = \beta * \beta_r$
  End A
End Program

where $\beta_0, \beta_r, \beta_f$ and I_0 correspond to the parameters of (Gold and Rangarajan 1996) and are application dependant. In our case, we used the values proposed in the original article. Function $ComputeQ$ computes optimizes the alignments and the point-to-point assignations, an outline of the procedure is given below:

Function $ComputeQ$ **input** P_h, p, a, w_1, Γ
returns Q^p_{a,w_1}
  For $\forall q \in 1..|\Gamma|, p \neq q$
    $P^{p,q}_f = P^p_h \cdot \left(P^q_h\right)^T$
    For $i = 1..|O^q|$
      For $b = 1..|O^p|, b \neq a$
        For $j = 1..|O^q|, i \neq j$
          $Q^p_{a,w_1} = Q^p_{a,w_1} + P^q_h[i,w_1]\cdot$
          $\cdot P^{p,q}_f[b.j]\cdot\left(\rho - \left\|\vec{x}_{ai} - \Phi^*_{ai}\vec{y}_{ij}\right\|^2_\Sigma\right)$
        End
      End
    End
  End
End Function

Taking into account our definition of outlier detection, we require to adapt the Sinkhorn normalization (Sinkhorn 1964) to consider them.

Recall first that the resulting Q^p_{a,w_1} could be negative values, however after the exponentiation all values become strictly positive and therefore we can assume the Sinkhorn normalization can be applied. In the normalization over matrix P_h, we keep in mind that outliers are special assignation that only satisfy one-way constraints, in this way we can easily consider several points as outliers. To this aim, we enhance each matrix P_h with an extra row and column, following a similar procedure than the slacks in (Gold and Rangarajan 1996). We initialize these extra row and column with the value of 1. We aim to detect outliers, that is points which have $R_{ai} < 0 \; \forall \; i$ or a. We know that $e^x \geq 1$ if $x \geq 0$, thus we expect points which have all possible assignations negative are assigned to this special row or column. Finally, when the Sinhorn method has finished the extra row and column are removed leading to the resulting matrices of global assignments P_h.

Note that now P_h cannot be theoretically considered a probability assignation matrix, due to $\sum_{a=1}^{op} P_h[a,i] \neq 1$, neither for rows nor for columns. However, we still can ensure that $\sum_{a=1}^{op} P_h[a,i] < 1$ and that each individual value is positive. So, what was a probability matrix P_h, now it can be assumed to be a fuzzy assignation matrix.

6 EVALUATION

We evaluate the effectiveness of the presented method in a series of group-wise image registration experiments. We use real images from the database in (Mikolajczyk, Tuytelaars et al. 2011). Feature points from each image have been extracted using the Harris operator (Harris and Stephens 1988). We use the following datasets: New York, Van Gogh and Asterix. Each dataset is composed by an ordered sequence of images from the same scene showing increasing levels of zoom or zoom plus rotation. Each test is performed on a group of N images. We compare the following four methods. (1) Pairwise ICP+RANSAC, which applies the well-known ensemble ICP+RANSAC between each pair of images. (2) Confident ICP+RANSAC, which computes the labellings between only the most similar pairs and infers the rest by composition (this method exploits the prior knowledge about the underlying order of the images). A very similar strategy is used in (Williams, Wilson et al. 1997). (3) Pairwise Labelling, which applies the proposed approach independently to each pair of images and (4) Group-wise Labelling, which applies the proposed approach jointly to all the images of the group. This method is the prime motivation of our work. When comparing the last two methods, it is our aim to elucidate the benefits of the group-wise approach vs. the pairwise one. All the methods have

Table 1: Results using New York. We have used 25 groups of N=4 images (i.e., results are averaged over 25 experiments).

	img[X]	img[X+1]	img[X+2]	img[X+3]
img[X]	-	24,7	127,24	199,26
	-	24,7	84,58	408,774
	-	**4,37**	131,84	161,02
	-	14,13	**16,34**	**20,88**
img[X+1]	82,21	-	44,91	108,57
	19,92	-	26,75	73,03
	4,57	-	4,37	131,83
	8,71	-	**3,45**	**7,77**
img[X+2]	170,74	23,78	-	27,4
	56,95	22,96	-	23,08
	118,27	4,56	-	**4,34**
	10,93	**3,37**	-	4,74
img[X+3]	245,95	83,3	25,4	-
	119,92	57,83	22,45	-
	174,6	118,25	**4,53**	-
	16,84	**8,55**	5,66	-

Table 2: Results using Van Gogh. We have used 14 groups of N=4 images (i.e., results are averaged over 14 experiments).

	img[X]	img[X+1]	img[X+2]	img[X+3]
img[X]	-	23,33	121,28	105,31
	-	23,33	91,62	318,97
	-	**0,85**	14,92	65,46
	-	1,44	**1,12**	**3,6**
img[X+1]	20,51	-	14,4	151,2
	13,33	-	15,39	281,13
	0,7	-	**0,66**	20,2
	1,27	-	0,69	**2,45**
img[X+2]	41,47	15,3	-	22,02
	28,86	14,15	-	17,77
	22,4	**0,52**	-	**0,52**
	0,68	0,54	-	1,61
img[X+3]	56,55	33,34	12,38	-
	132,53	49,48	11,72	-
	45,27	27,67	**0,42**	-
	1,72	**1,6**	1,17	-

Table 3: Results using Asterix. We have used 17 groups of N=4 images (i.e., results are averaged over 17 experiments).

	img[X]	img[X+1]	img[X+2]	img[X+3]
img[X]	-	134	71,1	104,8
	-	125,28	710,61	53,57
	-	**1,28**	9,88	34,65
	-	**1,28**	**1,42**	**3,44**
img[X+1]	23,2	-	53,9	70,7
	19,16	-	31,33	115,75
	1,05	-	**1,16**	14,01
	1,07	-	1,21	**2,47**
img[X+2]	58,2	36,7	-	37
	74,58	52,65	-	36,51
	14,59	**1,02**	-	3,5
	0,63	1,06	-	**3,01**
img[X+3]	1135,2	145,1	45,7	-
	50,36	74,84	129,69	-
	29,04	11,4	2,63	-
	1,78	**1,47**	**2,44**	-

been initialized with the results of the matching by correlation. Regardless the labellings are computed in either pair-wise or a group-wise fashion, results are evaluated in a pair-wise basis. We use the DLT algorithm (Kovesi 2009) to compute the homography corresponding to a given labelling between two images. Since ground truth homographies are available, we measure the accuracy through the mean projection error (MPE) in pixels.

Tables 1, 2 and 3 show the results of the New York, Van Gogh and Asterix datasets using groups of N=4 images. From top to bottom, each cell contains the MPE of Pair-wise ICP+RANSAC, Confident ICP+RANSAC, Pair-wise Labelling and Group-wise Labelling. Images are arranged in the rows and columns of the tables according to their logical order. The diagonal cells are empty since they correspond to self-labellings.

Analyzing the results, we see that the common labelling approach obtains usually the lowest mean projection error.

This fact is clear with distant images; see for instance row $img[X]$ and $img[X+3]$ where in all datasets the common labelling error is much lower with respect to all other methods. In some cases, with adjacent images the pair-wise labelling method obtains better labellings, e.g. row $img[X+3]$ and column $img[X+2]$ of Table 1. However, the difference between this method and the common labelling method is low, recall that the mean projection error is in pixels.

In addition to MPE, we show three concrete examples (Figs. 3, 4, 5, 6, 7 and 8) of labellings obtained with the pair-wise method and the common labelling method. Figs. 3 and 4 show an example over the Asterix dataset, Figs. 5 and 6 an example over New York dataset and finally Figs. 7 and 8 an example over the Van Gogh dataset. See how the method is able to remove incorrect matches, select better point matchings and increase the amount of point matches found. The first case is clearly seen in the Asterix example, the common labelling is able to detect that the points from the belly of Obellix do not correspond to the top letters. The second case is exemplified in the Van Gogh Example, the common labelling methods is able to correct several point matchings giving more than an acceptable result.

Figure 3: Concrete labelling example of Asterix dataset using obtained using pair-wise method.

Figure 4: Concrete labelling example of Asterix dataset using obtained using common labelling method.

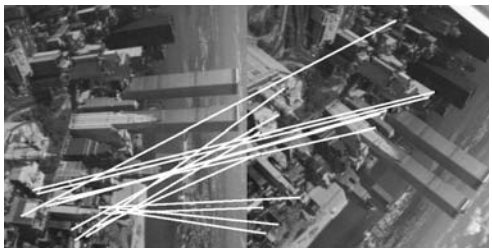

Figure 5: Concrete labelling example of New York dataset using obtained using pair-wise method.

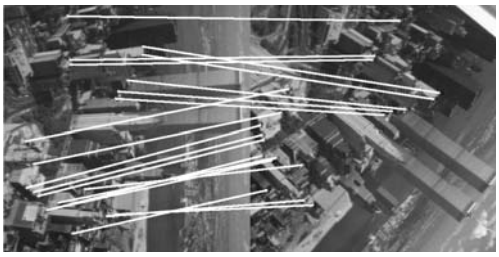

Figure 6: Concrete labelling example of New York dataset using obtained using common labelling method.

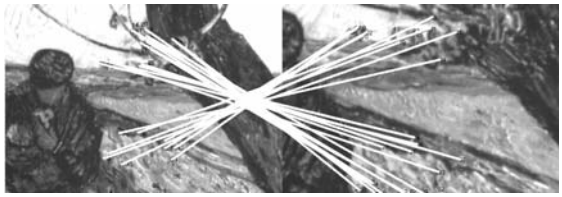

Figure 7: Concrete labelling example of Van Gogh dataset using obtained using pair-wise method.

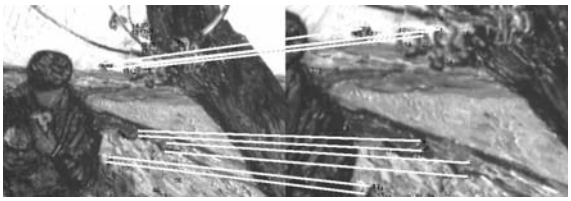

Figure 8: Concrete labelling example of Van Gogh dataset using obtained using common labelling method.

Finaly, the example over the New York dataset shows how the common labelling is able to match a greater amount of points with a better accuracy.

7 CONCLUSIONS

In this article, we have presented a group-wise method to compute sparse correspondences among a set of images. The main motivation is that pair-wise image labellings within a group can be significantly improved when solved jointly for all the members instead of independently for each pair. Moreover, the method can be used to compute pair-wise labelling, in this case the method considers jointly the labelling from image 1 to image 2 and vice versa. The method exploits relational geometrical information between pairs of points in an affine invariant way in order to compute pair-wise labelling compatibilities. Such geometrical compatibilities are used to feed a common labelling framework aimed at providing global consistency. Experiments show that the presented method improves considerably pair-wise labellings between distant images with respect to the other methods. Occasionally, this improvement is made at the cost of slightly penalizing the labellings between adjacent images.

ACKNOWLEDGEMENTS

This research is supported by “Consolider Ingenio 2010”: project CSD2007-00018, by the CICYT project DPI2010-17112 and by the Universitat Rovira I Virgili through a PhD research grant.

REFERENCES

Bonev, B., F. Escolano, et al. (2007). Constellations and the unsupervised learning of graphs. International conference on Graph-based representations in pattern recognition: 340-350.

Cootes, T., C. Twining, et al. (2010). "Computing Accurate Correspondences across Groups of Images " Pattern Analysis and Machine Intelligence 32(11): 1994-2005.

Fergus, R., P. Perona, et al. (2007). "Weakly Supervised Scale-Invariant Learning of Models for Visual Recognition." International Journal of Computer Vision 71 (3): 273-303.

Fischler, M. and R. Bolles (1981). "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography." Communications of the ACM 24(6): 381-395.

Gold, S. and A. Rangarajan (1996). "A Graduated Assignment Algorithm for Graph Matching." Transaction on Pattern Analysis and Machine Intelligence 18(4): 377-388.

Gold, S., A. Rangarajan, et al. (1998). "New algorithms for 2d and 3d point matchin." Pattern Recognition 31: 1019-1031.

Harris, C. and M. Stephens (1988). A Combined Corner and Edge Detection. The Fourth Alvey Vision Conference.

Horaud, R., F. Forbes, et al. (2011). "Rigid and articulated point registration with expectation conditional maximization." Pattern Analysis and Machine Intelligence 33: 587-602.

Hummel, R. and S. Zucker (1983). "On the foundations of relaxation labling processes." Pattern Analysis and Machine Intelligence 5(3): 267-287.

Jian, B. and B. Vemuri (2005). A robust algorithm for point set registration using mixture of gaussians. International Conference on Computer Vision.

Jian, B. and B. Vemuri (2011). "Robust point set registration using gaussian mixture models." Pattern Analysis and Machine Intelligence 33: 1633-1645.

Kovesi, P. (2009). "http://www.csse.uwa.edu.au/~pk/Research/MatlabFns/."

Mikolajczyk, K. and C. Schmid (2005). "A performance evaluation of local descriptors " Transaction on Pattern Analysis and Machine Intelligence 27(10): 1615-1630.

Mikolajczyk, K., T. Tuytelaars, et al. (2011). "http://www.featurespace.org/." Retrieved 23/02/2011.

Myronenko, A. and X. Song (2010). "Point Set Registration: Coherent Point Drift." Pattern Analysis and Machine Intelligence 32(12): 2262-2275.

Rangarajan, A., H. Chui, et al. (1997). The softassign procrustes matching algorithm. International Conference on Information Processing in Medical Imaging.

Rosenfeld, A., R. A. Hummel, et al. (1976). "Scene Labeling by Relaxation Operations." Transactions on Systems, Man, and Cybernetics 6: 420-443.

Sinkhorn, R. (1964). "A Relationship Between Arbitrary Positive Matrices and Doubly Stochastic Matrices." The Annals of Mathematical Statistics 35(2): 876-879.

Solé-Ribalta, A. and F. Serratosa (2010). Graduated assignment algorithm for finding the common labelling of a set of graphs. International conference on Structural, syntactic, and statistical pattern recognition 180-190.

Solé-Ribalta, A. and F. Serratosa (2011). "Models and algorithms for computing the common labelling of a set of attributed graphs." Computer Vision and Image Understanding 115(7): 929-945.

Wang, F., B. Vemuri, et al. (2008). "Simultaneous Nonrigid Registration of Multiple Point Sets and Atlas Construction " Pattern Analysis and Machine Intelligence 30(11): 2011-2022.

Williams, M., R. Wilson, et al. (1997). "Multiple Graph Matching with Bayesian Inference." Pattern Recognition Letters 18: 1275-1281.

Wong, A. and M. You (1985). "Entropy and Distance of Random Graphs with Application to Structural Pattern Recognition." Transaction on Pattern Analysis and Machine Intelligence PAMI-7(5): 599-609.

Zhang, Z. (1992). "Iterative Point Matching for Registration of Free-form Curves." International Journal of Computer Vision 13(2): 119-152.