UNIVERSITÀ CA' FOSCARI DI VENEZIA

DOTTORATO DI RICERCA IN INFORMATICA, 22° CICLO

(A.A. 2006/2007 – 2008/2009)

A Game-Theoretic Approach to Matching and Robust Inlier Selection

SETTORE SCIENTIFICO DISCIPLINARE DI AFFERENZA: INF/01

TESI DI DOTTORATO DI ANDREA ALBARELLI, 955379

TUTORE DEL DOTTORANDO

prof. Andrea Torsello

COORDINATORE DEL DOTTORATO

prof. Antonino Salibra

January, 2010

Author's Web Page: http://www.albarelli.it

Author's e-mail: albarelli@unive.it

Author's address:

Dipartimento di Informatica
Università Ca' Foscari di Venezia
Via Torino, 155
30172 Venezia Mestre – Italia
tel. +39 041 2348411
fax. +39 041 2348419
web: http://www.dsi.unive.it

Abstract

Many matching problems involve some sort of combinatorial optimization in order to find a solution that maximizes a non-convex global utility function. While in the most general context an exhaustive search of the solution space may be necessary, often the structure of the problem itself allows for some effective heuristics. For instance backtracking and reactive search algorithms perform well in the subgraph isomorphism problem, where the optimal solution offer a big attraction basin. Further, more constrained problems, such as bipartite matching or tree isomorphism have been proven to be solvable in polynomial time.

In this thesis we focus our attention on matching problems where some compatibility function can be defined over sets of two or more matched pairs. Specifically, we propose a flexible approach that exploits game theory in order to evolve an initial population of hypothesis to an evolutionary stable state where a smaller set of highly compatible matches survived. The rationale of this approach is two-fold. In fact, from a theoretical point of view it can be shown that in many problem formulations evolutionary stable states or Nash equilibria correspond to desirable configurations of the solution, such as maximal isomorphisms or optimal image or surface alignment. Moreover efficient algorithms exist to drive the evolutionary process and we will show with a very extensive set of experiments that even simple dynamics are able to lead the population to the optimal match. By applying our framework to different scenarios we will prove its effectiveness both in pairwise matching and in higher order problems, where different adaptations are proposed.

In addition to the described application to matching, the proposed approach can easily be used to solve parameter estimation problems. In particular it can be adapted to perform as a very robust inlier selector in all those problems where a level of compatibility with respect to some parameter configuration can be calculated over sets of two or more data points. The main difference between our approach and typical RANSAC-based solutions is that we do not need to rely on consensus checks over random picked local hypothesis; rather we are able to maintain a global consistency at every step of the evolution process, thus allowing for a more predictable final configuration that does not require any hard thresholding. We will show that this process is very robust and quite effective in a wide range of applications.

With respect to the experimental validation of the technique, we applied it to several different topics. We used it to match relational structures in an object retrieval and

tracking applications. Very good results have been obtained in point-pattern matching, segmentation matching and affine parameter estimation in general. Finally two applications in 3D reconstruction have been explored: the search for optimal symmetries in point clouds and the fine surface alignment of range images. The results obtained heretofore show both a great flexibility and robustness of the framework and stimulate additional work toward a further generalization of the approach and a widened application portfolio.

Sommario

Una fetta consistente dei problemi di matching richiede l'utilizzo di tecniche di ottimizzazione combinatoria per la ricerca di una soluzione che massimizzi globalmente una qualche funzione non convessa. Benché nel caso piú generale possa essere necessaria una ricerca esaustiva all'interno dello spazio delle soluzioni, spesso la stessa struttura del problema permette l'adozione di euristiche efficaci. Per esempio tecniche di backtracking o di reactive search possono dare risultati soddisfacenti nel problema dell'isomorfismo di sottografi, dove la soluzione ottimale presenta solitamente un bacino di attrazione di grandi dimensioni. Inoltre problemi dotati di maggiori vincoli, come il matching bipartito o l'isomorfismo di alberi, possono essere addirittura risolti in modo esatto in tempo polinomiale.

In questa tesi la nostra attenzione è rivolta alla classe di problemi di matching dove qualche funzione di compatibilit e definibile su un insieme di due o pi coppie corrispondenti. Nello specifico proponiamo un approccio flessibile che sfrutta la teoria dei giochi per permettere l'evoluzione di un'iniziale popolazione di ipotesi verso uno stato evolutivamente stabile dove un ristretto insieme di corrispondenze altamente compatibili è riuscito a sopravvivere. La motivazione che spinge ad adottare tale approccio è duplice. Infatti da un punto di vista teorico è possibile dimostrare che in molte formulazioni di problemi gli stati evolutivamente stabili o gli equilibri di Nash corrispondono a configurazioni desiderabili della soluzione, quali ad esempio isomorfismi massimali o allineamenti ottimali di superfici. Inoltre sono disponibili in letteratura molti algoritmi efficienti per guidare il processo evolutivo e, come mostreremo con un'estensiva copertura sperimentale, persino le dinamiche più semplici permettono di condurre la popolazione iniziale verso un match ottimale. Applicando il nostro framework a diverse tipologie di scenario mostreremo la sua efficacia sia in contesti di matching pairwise, sia coinvolgendo compatibilità di ordine superiore.

In aggiunta alle applicazioni al matching appena descritte, l'approccio proposto puó essere facilmente adottato per risolvere problemi di stima di parametri. In particolare puó essere utilizzato come robusto selezionatore di inlier in tutti quei problemi dove è possibile definire fra insiemi di due o piú punti dati un livello di compatibilitá relativo a qualche configurazione di parametri. La differenza principale fra il nostro approccio e le tipiche soluzioni basate su schemi di tipo RANSAC è che nel nostro caso non ci affidiamo a controlli di consenso a posteriori a partire da ipotesi iniziali scelte casualmente, ma piuttosto siamo in grado di mantenere una consistenza globale ad ogni passo del processo evolu-

tivo, permettendo cosí di raggiungere in maniera piú deterministica una configurazione finale ottimale senza richiedere nessun tipo di soglia prefissata. Mostreremo che questo processo é robusto ed efficace in un grande numero di applicazioni.

Relativamente alla validazione sperimentale la tecnica proposta è stata applicata a diversi ambiti. Ad esempio 'e stata utilizzata per il matching di strutture relazionali nell'object retrieval e nel tracking. Ottimi risultati sono stati ottenuti nel point-pattern matching, nella segmentazione e nella stima di trasformazioni affini in generale. Infine abbiamo esplorato due applicazioni nel campo della ricostruzione 3D: la ricerca di simmetrie all'interno di nuvole di punti a l'allineamento fine di superfici provenienti da range image. I risultati ottenuti fino a questo punto hannod mostrato una grande flessibilità e robustezza del framework, incoraggiando ulteriore ricerca sulla sua generalizzazione ed applicazione ad un più vasto insieme di problemi.

Contents

List of Figures

Preface

This thesis summarizes the work done during my post graduate studies with the intent to offer to the reader a coherent and comprehensive coverage of the different topics. While the main focus of my thesis became more and more apparent with the progress of my research, many paths have been explored along the journey. Each one of them has been useful under multiple points of view, giving to me the opportunity to study in deep new problems, develop original ideas and compare them with colleagues and peers. The first problem I tackled was the search for subgraph isomorphisms in attributed graphs. While not solved directly with a game-theoretic approach I casted the problem in a clique search, which gave me the first idea about modeling associations as items related by a level of compatibility. This work resulted in a paper published in [10]. Afterwards I extended my approach to the tracking of articulated object by matching their structural representation. This technique lead to a collaboration with the University of Modena and Reggio Emilia and to a publication in [8]. The graph matching topic was further developed by defining a novel type of association hypergraph used for parameter estimation published in [7] and a consensus graph that can be used for finding symmetry planes in three-dimensional point clouds [1]. This latter work was prompted by my growing interest in 3D data reconstruction and analysis. In this field I introduced a novel and very effective coarse and fine registration method [3] and a fast compound phase coding technique [2] developed during the engineering of a full fledged structured light based 3D scanner. In addition a robust inlier selection technique for bundle adjustment is presented in [5]. Finally, my work on game-theoretic matching culminated in a more mature framework presented, along with some interesting applications, in [6]. This framework is indeed the natural meeting point of most of the research threads I followed in the last three years, and as such represents the logical starting point for this thesis.

Published Papers

[1] ALBARELLI, A., PELILLO, M., AND VIVIANI, S. Consensus graphs for symmetry plane estimation. In *SSPR & SPR '08: Proceedings of the 2008 Joint IAPR International Workshop on Structural, Syntactic, and Statistical Pattern Recognition* (Berlin, Heidelberg, 2008), Springer-Verlag, pp. 197–206.

[2] ALBARELLI, A., RODOLÁ, E., BULÓ, S. R., AND TORSELLO, A. Fast 3d surface reconstruction by unambiguous compound phase coding. In *IEEE International Workshop on 3D Digital Imaging and Modeling (3DIM2009)* (2009).

[3] ALBARELLI, A., RODOLÀ, E., AND TORSELLO, A. A game-theoretic approach to fine surface registration without initial motion estimation. In *Submitted to CVPR 2010* (2010).

[4] ALBARELLI, A., RODOLÀ, E., AND TORSELLO, A. Robust figure extraction on textured background: a game-theoretic approach. In *Submitted to ICPR 2010* (2010).

[5] ALBARELLI, A., RODOLÀ, E., AND TORSELLO, A. Robust game-theoretic inlier selection for bundle adjustment. In *Submitted to 3DPVT 2010* (2010).

[6] ALBARELLI, A., TORSELLO, A., AND BULÓ, S. R. Matching as a non-cooperative game. In *International Conference on Computer Vision (ICCV2009)* (2009).

[7] BULÒ, S. R., ALBARELLI, A., TORSELLO, A., AND PELILLO, M. A hypergraph-based approach to affine parameters estimation. In *ICPR* (2008), IEEE.

[8] GUALDI, G., ALBARELLI, A., PRATI, A., TORSELLO, A., PELILLO, M., AND CUCCHIARA, R. Using dominant sets for object tracking with freely moving camera. In *Int. Workshop on Video Surveillance* (2008).

[9] RODOLÀ, E., ALBARELLI, A., AND TORSELLO, A. A game-theoretic approach to robust selection of multi-view point correspondence. In *Submitted to ICPR 2010* (2010).

[10] TORSELLO, A., ALBARELLI, A., AND PELILLO, M. Matching relational structures using the edge-association graph. *International Conference on Image Analysis and Processing* (2007), 775–780.

Notations

General

$\mathbb{R}$	set of reals
$\mathbb{R}_+$	set of nonnegative reals
$\|S\|$	cardinality of the set S
$\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$,...	column vectors in $\mathbb{R}^n$
x_i	ith element of vector $\mathbf{x}$
$\mathbf{0}$	column vector of all 0s
$\mathbf{e}$	column vector of all 1s
$\mathbf{e}^i$	ith column of the identity matrix
$A = (a_{ij})$	matrix in $\mathbb{R}^{n\times n}$ with (i,j)th element a_{ij}
I	identity matrix
$det(A)$	determinant of matrix A
$\binom{V}{k}$	set of k-subsets of V

Game-theory

$S = \{1,\ldots,n\}$	set of n strategies usually denoted with S
Δ^n	standard simplex in n dimensions
$\pi(\mathbf{y}\|\mathbf{x})$	$\mathbf{y}^T A\mathbf{x}$, average payoff of $\mathbf{y}$ against $\mathbf{x}$
$\pi(\mathbf{x})$	$\mathbf{x}^T A\mathbf{x}$, average population payoff
$\sigma(\mathbf{x})$	support of vector $\mathbf{x}$
$\beta(\mathbf{x})$	set of best replies to strategy $\mathbf{x}$
$\mathbf{x}^{(t)}$	population state at time t
$\dot{\mathbf{x}}^{(t)}$	time derivative of the population state at time t

1

Introduction

At the very heart of a large number of model fitting and parameter estimation problems a matching challenge lies. In fact, when trying to justify some observed output with respect to input data points, two class of hurdles arise. The first and most apparent one is about the model itself and its parameters. An appropriate mathematical description of the phenomenon must be formulated and an optimization strategy must be chosen. Unfortunately, even with a perfect model and an optimization technique, guaranteed to converge to a global minimum, an often overlooked problem stands in the way to parameter estimation. This problem is the elimination of the outliers and the correct matching between data sets. In model-fitting problems linear and non-linear optimizations are widely used in order to find a least square interpolation of the data. This technique is quite standard and works very well when the measurements are subject to small noise. By converse, when outliers are present they have significant influence since their distance from the model is inherently magnified in the error computation (see Fig 1.1). Of course many outlier detection techniques have been proposed in literature. Some approaches aim to the iterative refinement of the fitted model by removing at each step data that exhibit a too large error, others apply a transformation from the data to the parameters space and search for global or local maxima. All of these technique have to deal with several quandaries, such as the compromise between precision and recall or the choice of appropriate thresholds to match the data point to the fitted model.

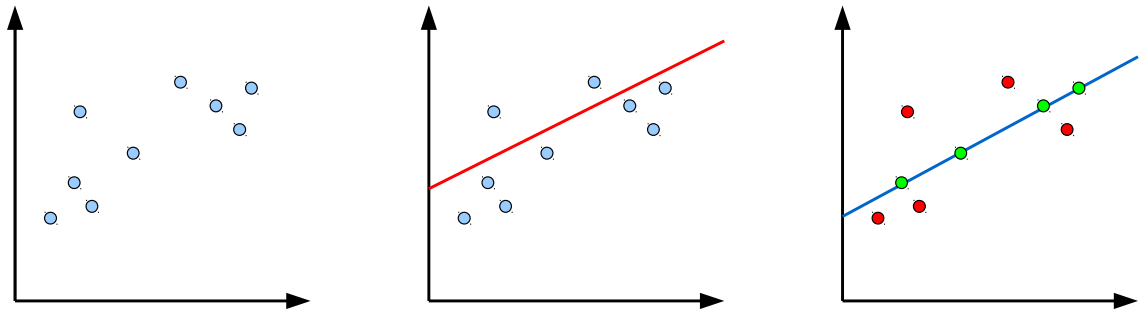


Figure 1.1: Example of failing least squares fitting. Outliers lead to wrong fitting (second image). Accurate inliers selection allows to find the ground truth model (third image).

Another class of parameter estimation problems involves the matching between two sets of data points with the goal of recognize a pattern or estimate an unknown transformation between them. Repeatable feature points can be used to find an object in clutter or to estimate an affine transformation between a pair of images. In the latter case, the presence of outlier can lead to inaccurate estimations or to completely wrong matches, depending on the optimization technique chosen. For instance, RANSAC-based techniques can easily fail when several correspondences configurations with high consensus exist, in addition such methods are strongly dependent on the number of iteration allowed and the threshold fixed for a valid vote (see Fig 1.2).

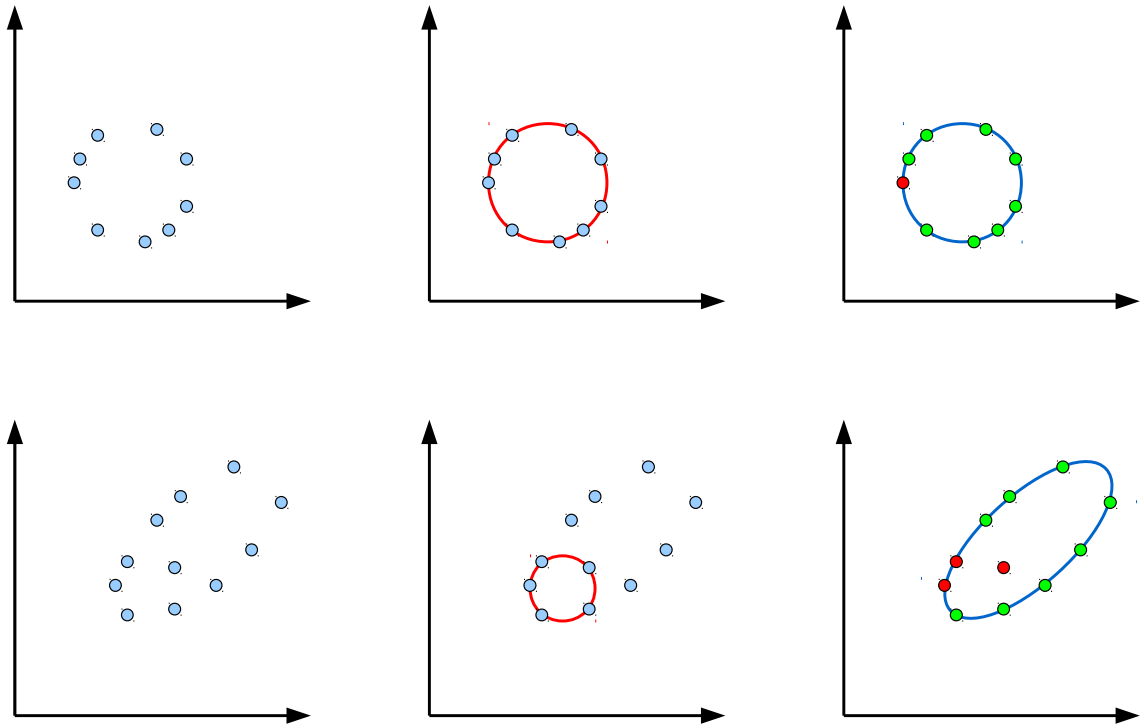


Figure 1.2: Example of failing RANSAC-based matching. Outliers lead to a wrong match that exhibits high consensus (second column) and thus to the estimation of the wrong transformation. The correct matching has slightly higher consensus (third column).

In chapter 2 we will review most of the currently adopted techniques for solving matching and inlier selection problems. For each technique strengths and weakness will be highlighted. In the subsequent chapters a novel game-theoretic based approach will be introduced and its theoretical properties and empirical effectiveness will be discussed. Depending on the application context, comparisons with the most appropriate techniques will be made.

Globally Coherent Matches by Population Evolution

The main idea developed throughout this thesis is quite simple. In general, when matching two sets of item, both local and global consistency checks can be performed. A local check is an unary function over the domain of the single matches. This can be, for instance, a binary function that states if the match is allowed, or a positive real value that express some degree of compatibility between the two mated points. By converse, global checks are evaluated over the entire matched subset and usually represent some utility function to maximize. Most matching techniques use local checks to obtain an initial set of viable matches that will be used as a base for subsequent optimization. Our idea is to operate an early global selection, by adding more-than-unary local checks and exploiting the transitive nature of the concept of compatibility. We assume that two conditions are satisfied in the tackled problem: that some measure of compatibility is defined over sets of two or more matches and that some technique allows for a selection of a set of matches with high mutual compatibility. The measure of compatibility depends on the scenario. As we will show in the following chapters, different compatibility definitions can be used to solve a wide range of problems. Also the selection of matches with high mutual compatibility can be done with many approaches. We choose to explore the use of game theory for this purpose. Specifically we model each possible match candidate as a strategy in a non-cooperative game and assign to each pair of strategies a payoff that is proportional to the measure of compatibility between such matches. The rationale of this choice is that by letting an initial population of strategies evolve to a stable state we are aiming to obtain a configuration of surviving strategies with an high mutual payoff and thus globally coherent. The complete approach is discussed for the first time in Chapter 3 with application to pattern matching and further expanded in Chapter 5 with some applications to 3D reconstruction.

Higher-order Problems and Structural Matching

Our game-theoretic matching framework can be easily adapted to be used in applications where it is not possible or useful to define a compatibility functions between pairs of matches. For instance we could use our approach to replace Hough Transform for searching collinear cluster of points. Unfortunately, since any pair of points is collinear in the plane, by defining a compatibility measure between two points we would end up by selecting all the data point by means of transitive closure. This can be solved, for instance, by using a compatibility measure defined over pairs of couple of points. A possible method could be to model couple of data points as strategies and assign to each pair of strategies a compatibility proportional to collinearity of the lines defined by each strategy. This approach allows to handle all the problems where is necessary to model strategies with more of one data point. In general this is the case when we want to select a set of inliers with respect to a property that can be defined over n points and verified over at maximum $2n$ points. In this condition the transitive closure operated by the evolu-

tionary dynamics performs very well. While this approach works (as it will be shown in chapter 4), it has the drawback that the number of strategies grows polynomially with the cardinality of the property, thus is not very useful for practical purposes when n is greater than 2 or 3. These cases can be handled with a little different approach: strategies can still be modeled to single data points and compatibilities alone can be calculated over sets of strategies. Specifically, when we deal with a property that can be verified over k points, we can view strategies (and thus data points) as vertices in a $k - hypergraph$ where each subset of k vertices is connected by a hyper-edge if and only if the property is verified by them. In this context our notion of set with high mutual compatibility corresponds to a hyper-clique (i.e. a completely connected subgraph) of maximum cardinality. We will show how to use some effective heuristics in order to search for such a set.

Finally, in some problems, we need to preserve some structure or relational constraints. This is the case, for instance, when matching attributed graphs built on images of objects or articulated figures for matching or tracking purposes. In our framework structure can be enforced in several ways. One-to-one versus many-to-many constraints can be encoded directly in the payoff matrix of the modeled strategies by assigning a value of zero between strategies that share the same source or destination data point. More sophisticated structural constraints can be encoded by means of an auxiliary association graph that has edges connecting only pairs of strategies that preserve the desired structural properties. Several problems where higher order relations and structural preservation are needed will be covered in Chapter 4.

Matching applications in 3D reconstruction

Given the flexibility of out matching approach we decided to test it with some problems in the field of 3D reconstruction. Specifically we used it to find symmetry planes in point clouds and to match surfaces produced with a structured light 3D scanner. While many techniques exist for symmetry plane estimation we proposed a novel approach that models potentially symmetric points as strategies and assigns to pairs of strategies a payoff proportional to the compatibility of the respective symmetry planes. Comparisons with other well known approaches show measurable improvements in accuracy.

Further we used our game theoretic matching to solve at one the coarse and fine range-image alignment problems. Image based 3D scanners are only able to produce partial view of an object, primarily because of the injective nature of the structured light source and image sensors. In order to obtain a full reconstruction of the subject it is necessary to align with high accuracy several of these partial views, commonly called range-images. This is usually a two-step process: a coarse registration, based on high level features, is first obtained; then a refinement in performed by iteratively match nearest compatible surface points. The dichotomy between coarse and fine registration is justified by the need to operate with low selectivity when, in absence of an initial estimation, spatial information is useless to evaluate the compatibility between points. Since our approach allows for a high selectivity level even when few information is carried by the features itself, it can

be used to obtain an accurate fine registration without any previous motion estimation. In addition, by contrast to iterative closest points techniques, is not affected at all by local minima.

Finally, during our work in 3D reconstruction, we had the opportunity to build an in-house scanning system and to experiment with a novel pattern strategy for fast and unambiguous compound phase shift coding.

2

Matching, Parameter Estimation and Inlier Selection

In this chapter we present an extensive review of relevant matching and inlier selection techniques found in literature. While we try to offer a comprehensive view, we still keep the main focus on approaches that have goals similar to the methods discussed in this thesis. In particular we discuss all the techniques that will be used to evaluate the effectiveness of our approach in the following chapters. Where possible, appropriate pointers to specialized reviews are given.

2.1 Point Pattern Matching

Point pattern matching is a broadly used technique in data analysis. Specifically, in the computer vision field, it is a valuable tool for image registration, object recognition and matching of corresponding sets of object points in stereo vision. In fact, in most situations, it is not feasible or efficient to deal with the whole collection of pixel contained in a raster image. It is indeed preferred to extract from images a reduced set of salient or distinctive features to increase the significance of the information considered and to decrease the computation effort needed. This is also true for surfaces or solid object, where 3D interest points can be detected instead of 2D image features, and in general point patterns can always be obtained from any source of N-dimensional data. When searching for a correspondence between two point patterns a point pattern matching (PPM) algorithm is used. In general the goal of such technique is to find an optimal transformation such that a distance measure for the alignment of the two point patterns under this transformation is minimized. Roughly PPM problems can be categorized with respect to two characteristics: the completeness of the two sets to be matched and the availability of a known labelling for the data points. Complete matching requires that the two point patterns have the same number of points and there exists a one-to-one correspondence mapping between the two point sets. In incomplete matching, no such one-to-one correspondence mapping is constructed due to the missing and spurious points. Instead, incomplete matching seeks for a mapping between subsets of the point patterns. Both complete and incomplete matching can occur with or without the availability of point labels. When such a labelling

exists, it can be used as an a priori information that is additional to the point coordinates. Examples of such information are color, intensity, scale or any kind of data vector that is associated to the points and is somewhat repeatable under several observations. Clearly, incomplete matching is harder than complete matching and unlabeled matching is more complex than labeled matching. In this thesis we will deal with incomplete point pattern matching where labelling is sometimes available, but it is always used for reducing the number of initial feasible mates rather than in the matching process itself.

Since in general PPM problems sets are incomplete and noise is present in both position measurements and labels, a key problem to solve the matching (even in an exhaustive way) is to define an appropriate distance measure between data points with respect to a candidate transformation. Such a distance should take in account both the location error between each corresponding points and the total number of pairs matched (since it is obviously advisable to encourage matches that involves many points). Two distance measures are usually applied for these purposes. The first is the *Agrawal's heuristic dissimilarity (AHD)* measure[11]. The AHD is based on the sum of the squared euclidean distances between pairs of matched points, which is normalized by taking in account the ratio between the number of matching points and the minimum cardinality of the two matched sets (in order to avoid small sets of correspondences). Another distance measure is the *Partial Hausdorff distance* (PHD) proposed in [82], where an efficient method to compute the Hausdorff distance between two sets is introduced.

Since an exhaustive search in the parameter space with the goal of minimizing these (or other) distance measures is not feasible, many heuristics and approximated approaches have been proposed for various PPM problems.

Clustering methods calculate the transformation parameters for all combinations of point pairs from both patterns and increase the merit of the corresponding cell in an accumulator matrix[51, 169, 184, 172, 81, 170]. The clusters in the accumulator matrix are then detected, the cluster with the maximum merit corresponds to the optimal parameters. The clustering methods are computationally intensive due to the large number of combinations of point pairs and the dimensionality of the parameter space. Their main weakness is related to the need for a suitable quantization of the parameter space: in fact a too coarse quantization can lead to poor precision, while a too fine one can incur in poorly populated bins and thus to the inability to find a representative cluster.

Parameter decomposition methods divide the parameter estimation process into multiple phases. At the first phase, a selected parameter is estimated based on the domain knowledge such as the geometric invariant constraints. Then, at each of the following phases, one or more of the remaining parameters are estimated by referring to those parameters values previously determined, hence, the number of possible combinations between values of separate parameters is greatly reduced [72, 121]. However, the inaccuracy level of parameter estimation could be increased by propagation through various phases.

Relaxation methods iteratively update merit score of each point correspondence from both patterns given the merit scores of the other interacting point correspondences. The interacting point mappings are those that are mutually constrained for matching. The algorithm converges when those merit values become consistent (or hardly changed) and

the point mappings with the maximum merits are considered as the true transformation point correspondence [132, 120, 157].

Mount et al. [118] proposed a geometric branch-and-bound search of the transformation space and used the point alignment information to bound the search. They specify an approximation factor to guarantee the accuracy of the final match and use point alignments when a significant number of point correspondences can be inferred to accelerate the search. The robustness of the algorithm has been demonstrated on registration of real satellite images.

Carcassoni and Hancock [36] applied the spectral graph theory to compute the point pattern correspondence. The global structural properties of the point pattern are ascribed by the eigenvalues and eigenvectors of the proximity weighting matrix. The influence of the contamination and drop-out in the point pattern is discounted via the EM algorithm so the accuracy of the matching is increased.

The search space of point mappings between two patterns can also be explored by genetic algorithms. The chromosomes encoding instances of point correspondences improve their fitness during evolution by use of three genetic operators: selection, crossover, and mutation. The fitness of the chromosome is defined as the inverse of the AHD function [18]. Zhang et al. [188] use the reference triplet points as the chromosome representation and thus significantly reduce the search space.

In [148] simulated annealing technique is applied to the point pattern matching problems. The identification of point correspondences between two point sets is mathematically formulated as energy minimization. The matching error corresponding to the current configuration of point correspondences is treated as the energy of that configuration. The configuration is iteratively rearranged to reach thermal equilibrium at various temperature levels and finally converges to an optimum as the system is frozen.

Finally, in [182], a particle swarm optimization algorithm is proposed. Following the optimization technique proposed in [91], the set of transformation parameters is encoded as a real-valued vector called particle. A swarm of particles are initiated at random and fly through the transformation space for targeting the optimal transformation.

2.2 Graph Matching

The characterizing goal of each graph matching problem formulation is the preservation of structural relations between corresponding nodes. Specifically, a match is considered consistent only if two source nodes connected by an edge are assigned to a pair of adjacent destination points and vice versa. Of course, depending on the problem formulation, the lack of consistency can be unacceptable or it can simply induce a penalty in the match score. While some instances of the graph matching problem have been shown to be solvable in polynomial time (for instance Bipartite Matching), in general no exact algorithms are available to tackle efficiently graph or subgraph matching. In fact all the commonly used exact approaches adopt expensive deterministic search techniques, usually based on backtracking [166, 113, 99, 124]. Polynomial convergence times are only achievable by

applying heuristics to the problem, which in turn allow to obtain approximate solutions. Given the importance of the graph matching problem and its variegation, many different approaches have been proposed over the last decades. Of course each technique differs from the others in term of conditions of applicability, expected convergence time and quality of the solutions found.

In [66] Gold e Rangarajan propose a weighted graph matching algorithm that is based on non-linear optimization. This method, named *Graduated Assignment*, exploits three main features: the enforcement of bidirectional constraints in the assignments, obtained by means of a softassign schema [122], the ability to avoid local minima, thanks to graduated non-convexity, and the use of sparsity to enhance convergence speed. We will propose an novel method for weighted graph matching in Section 4.1 and our results will be compared with those obtained with Graduated Assignment.

Many approaches found in literature are based on the application of probability theory and stochastic techniques. The first attempts in this sense are found in [76, 95] where an iterative approach based on probabilistic relaxation is used to match graphs with attributes on edges that are subject to Gaussian noise. More recently this approach has been augmented by using the Bayesian framework and it has been extended to deal with graphs that exhibit also attributes on nodes [43, 175]. In [174] Williams makes a comparative study of different deterministic discrete-space search strategies based on the Bayesian consistence measure proposed in [175]. In this study is also proposed the use of Tabu search techniques [134, 65, 63, 64] for the graph matching problem. Finally the framework has been extended further to adapt to the matching of relational structures based on hierarchical models [176]. All of these techniques are based on the iterative enhancement of an initial guess, thus them are strongly dependent on the method used to select this first estimate and on its quality. For this reason these approaches are only useful when the specific formulation of the matching problem allows to easily to detect a reasonable initial solutions, as they usually fail when starting from arbitrary points.

Another well known probabilistic approach is to adopt the EM (Expectation Maximization) algorithm, commonly used to make maximum likelihood estimation of parameters in probabilistic models. In literature we find several applications of this approach. For instance in [52, 59] the EM technique is applied to two different graph matching problem formulations. An algorithm that focuses on the matching of relational structures without attributes is proposed in [106] and extended in [107]. In [92] Expectation Maximization is applied to the recognition of handwritten characters that are represented as hierarchical graphs.

Other examples of techniques for the matching of simple or attributed graphs with probabilistic approaches can be found in the field of face recognition in [79], of biometric authentication (by means of probabilistic relaxation) in [54] e [110], and of the prediction of the creation of disulfide bridges in protein folding [56].

In [167] Umeyama propose a weighted graph matching technique that avoids combinatorial optimization and tries to exploit an analytical approach. The problem is solved efficiently by means of the spectral decomposition of the adjacency matrix, when dealing with directed graphs, and of an Hermitian matrix derived from the adjacency matrix, in

the case of undirected graphs. The experimental validation of the technique shows that it works well provided that the two graphs are similar enough. In addition, another limitation of this approach is that it is inherently able to work only with square matrices, thus it can only be used with graphs of the same size.

In [15] Almohamad introduces a graph matching approach based on linear optimization. At first, the maximum weight match problem is casted into a quadratic optimization problem. The latter is in turn reformulated as a linear optimization problem, which is then solved with the simplex algorithm. Once an optimal solution for the linear optimization is found, the solution is discretized by projecting its values in $\{0, 1\}$ through the Hungarian method, a well known combinatorial algorithm for the assignment problem proposed by Kuhn [97]. The overall approach exhibits a polynomial complexity, but unfortunately its order is quite high. In Fact it is $O(n^6 L)$ where n is the size of the graphs to be matched and L the size of the linear optimization problem. The algorithm is experimentally validated versus a technique based on symmetric polynomial transforms (SPT) [14] introduced by the same author, and versus the spectral decomposition technique suggested by Umeyama. In both cases the Almohamad approach is shown to found matches of higher weight.

2.3 RANSAC-based Techniques

The most widely used inlier selection technique is probably the Random Sample Consensus (RANSAC) [60]. RANSAC allow for a simultaneous estimation of the parameters of a model and segregation of outliers from inliers. Its main advantage is the ability to be robust with respect to a very large number of outliers. The general RANSAC framework can be adapted in any scenario where a small subset of data can be used to estimate a model which in turn can validate the remaining data. In fact the algorithm operate iteratively by randomly selecting a set of data points and fitting model parameters to them. Subsequently all the remaining data points are evaluated with respect to the estimated model and the number of points that are found to agree with it is calculated. The agreement of a point with the model (i.e. its consensus) can be evaluated in different ways, depending on the specific application. By repeating the selection and vote counting step many times the model with most consensus emerge. For instance when using RANSAC for line fitting, sets of 2 points are selected at random from the data and the number of points with a distance from the fitted line under a defined threshold are counted. After a predefined number of iterations the line which obtains the largest number of votes is considered correct. Of course, once a proper set of inlier is selected, a refinement technique, such as least square optimization, can be used to enhance the solution. While this technique works very well in general, it has at least three main drawbacks. The first is that when the number of outliers increases many iterations are needed to find a proper model candidate. This is specially true when the cardinality of the set needed for the estimation of the model parameters is large, as the probability of finding only inliers decreases exponentially with the cardinality of the set itself. The second drawback is that to evaluate the consensus of the data with respect to the model a threshold is needed in most situations.

This is unfortunate, as a too small threshold can lead to the impossibility to find a solution with a large enough consensus, while a too high one can result in a too much permissive selection. Finally it should be noted that, being an optimization technique based on random selection, RANSAC does not give any guarantee of convergence. In order to enhance the speed of the basic RANSAC scheme several optimizations have been proposed. We can classify such optimization int two categories: techniques that optimize the model verification stage and approaches that avoid to use uniform sampling in the parameters estimation step.

Matas and Chum [38] suggest a pre-evaluation step which attempts to filter out immediately wrong hypotheses. This is done by performing model verification on a small subset of data points: if this first test fails the evaluation stops and the hypothesis is discarded. Experimental validation has shown that the optimal cardinality of the first-stage verification set is 1. Of course this approach can lead to false negatives, nevertheless the reduced overall verification time obtained allows for a global reduction of the convergence time, despite the higher number of hypotheses to be tested.

The idea of early termination of bad model estimation was further extended by Capel in [35]. Given a randomly selected subset of points, the number of inliers in this subset follows a hyper-geometric distribution. Given the current best hypothesis with N inliers, each new hypothesis is first partially evaluated against a subset of the data points and the test is completed only if the probability of the total number of inlier to be more than N is above a fixed threshold.

More recently, Matas and Chum described an optimal randomized model verification strategy [112, 46] based on Walds theory of sequential decision making. The evaluation step is cast into an optimization problem which aims to decide whether a model is good or bad, while simultaneously minimizing the number of verifications performed per model.

The second category of ransac optimizations try to exploit a priori information to avoid to resort to the uniform sampling the input data set. The availability and the quality of such a priori information of course depends on the specific application context. For instance, correspondences between two or more images are obtained by the use of a local matching algorithm. A similarity function is evaluated over a number of points, and subsequently thresholded to obtain a set of tentative correspondences. Based on the assumption that points with high similarity are more likely to be inliers than points with low similarity, it may be possible to generate better hypotheses by sampling from a reduced set of points with high similarity scores. This fact is exploited in PROSAC [45], where the correspondences are ordered based on their similarity scores, and progressively larger subsets of tentative correspondences are used to generate hypotheses.

A similar approach to PROSAC was proposed earlier by Tordoff and Murray in [158], where the Maximum Likelihood Estimation Sample Consensus (MLESAC) algorithm [159] was combined with non-uniform sampling of correspondences. MLESAC algorithm is a generalization of RANSAC, which adopts the same sampling strategy but attempts to maximize the likelihood of the solution, as opposed to the number of inliers. While MLESAC assumes a uniform prior for the validity of a match, the guided-sampling approach of [158] uses the quality function of the feature matching algorithm to derive

probabilities of match validity. These are then incorporated as priors in MLESAC.

One of the assumptions inherent in the standard termination criterion of RANSAC is that a model computed from an uncontaminated sample is consistent with all inliers. In practice, this is often not the case, particularly when the data points are noisy. Chum et al. [47] define a locally optimized RANSAC variant to deal with this issue. By observing that a good model tends to find a significant fraction of the inliers, an inner RANSAC strategy is devised where a constant number of hypotheses are generated using only the set of inliers to the current best model. Since inner RANSAC operates on a set of inliers, it is not essential that hypotheses are generated from minimal subsets. In addition to providing a more robust fit, the inner RANSAC technique has the effect of improving the consensus score more rapidly than standard RANSAC, which causes the termination criterion to be met earlier.

2.4 Matching and Registration in 3D reconstruction

The registration of 3D surfaces coming from structured light scanners can be considered a specialized matching problem. Since in this thesis this problem will be addressed by means of our generalized game-theoretic matching framework, we provide here some introductory references related to the state of the art about this topic. A more detailed review will be presented in Chapter 5.

Surface registration is about finding an optimal alignment between two partially overlapping meshes captured with a 3D scanner or obtained with image based stereo reconstruction systems. In literature two different field of research are currently active: namely coarse and fine registration techniques. Fine registration algorithms exploit an initial guess in order to constrain the search area for compatible mates and minimize the risk of selecting outliers. By converse, coarse techniques usually adopt a mating strategy based on the similarity between surface-point descriptors.

Most fine alignment methods are modifications to the original ICP proposed by Zhang [189] and Besl and McKay [26]. These variants generally differ in the strategies used to sample points from the surfaces, reject incompatible pairs, or measure error. In general, the precision and convergence speed of these techniques is very sensitive to the fine-tuning of the model parameters. In order to overcome these limitations some variants avoid hard culling by assigning a probability to each candidate pair by mean of evolutionary techniques [101] or Expectation Maximization [70]. Other fine registration methods include the well-known method by Chen [41] and signed distance fields matching [111].

Coarse registration techniques can be roughly classified in methods that exploit some global property of the surface, such as PCA [48] or Algebraic Surface Model [156], and methods that use some 3D feature descriptor to find plausible candidates pairs over the model and data surfaces. Global techniques are generally very sensitive to occlusion. Feature-based approaches are more precise and can align surfaces that exhibit only partial overlap. Nevertheless, the unavoidable localization error of the feature points prevents them from obtaining accuracies on par with fine registration methods. Among the most

successful descriptors are Point Signatures [44] and Spin Images [89]. A completely different coarse registration approach is the RANSAC-based DARCES [39], which is based on the random extraction of sets of mates from the surfaces and their validation based on the accuracy of the estimated transformation. A recent and extensive review of all the different methods can be found in [138].

2.5 Structured Light Reconstruction

The very last chapter of this thesis is devoted to the study of a novel pattern coding strategy for trinagulation based on structured light. The main challenge for this kind of surface reconstruction techniques is the assignment of correspondences between features observed by two or more different points of view. Those correspondences can be obtained in several ways, ranging from feature extraction to the placements of artificial markers in the scene. Our attention is towards techniques that exploit the projection of some known pattern overt the objects to be acquired.

Among these, time-multiplexing strategies such as n-ary and Gray codes, as well as hybrid approaches, are by far the most utilized [88]. These techniques generate unique codes along each scanline, but at the same time are limited by their low resolution due to the inherently discrete nature of the coding. Phase shifting methods, on the other hand, yield higher resolutions since they are based on the projection of periodic patterns with a given spatial period. Each projected pattern is obtained by spatially shifting the preceding one of a fraction of the period, and then captured by one or more cameras. The images are then elaborated and the phase information at each pixel determined by means of M-step relationships [153]. A major drawback is that, in its basic formulation, phase shifted structured light renders only relative phase values and thus it is ambiguous. However, when both an extended measuring range and a high resolution are required, a combined approach proves to be very powerful. The integration of Gray code and phase shift brings together the advantages of both, providing disambiguation and high resolution, but the number of patterns to be projected increases considerably, and each strategy introduces a source of error [100].

Other high resolution shape measurement systems include optical profilometers. Non-contact phase profilometry techniques relate each surface point to three coordinates in a frame having the z axis orthogonal to a reference plane, which then represents the reference for the measured height [147, 178]. In classical phase measurement profilometry, gratings or sinusoidal patterns are projected and shifted first onto the plane and then over the object to be measured. Phase information from the deformed fringe pattern is then extracted by means of various techniques. Other, more effective profilometry techniques include the well-known Fourier Transform method [154] and other interesting derivatives [183, 73]. Fourier-based profilometry can require as few as one or two frames for depth estimation, which makes real-time reconstruction possible. Nevertheless, in profilometric methods phase variation caused by height modulation must be limited: ambiguity of the phase limits the measurement range, allowing for gauging of smooth-shaped objects only.

Moreover, noise from camera, distortion of lens, difficulties of calibration, aliasing and imperfectness of the projecting unit influence the precision of most of these techniques [40, 149].

3

Matching as a Non-Cooperative Game

In this chapter we offer a game-theoretic perspective for the all-pervasive matching problem in computer vision. Specifically, we formulate the matching problem as a (population) non-cooperative game where the potential associations between the items to be matched correspond to (pure) strategies, while payoffs reflect the degree of compatibility between competing hypotheses. Within this formulation, the solutions of the matching problem correspond to evolutionary stable states (ESS's), a robust population-based generalization of the notion of a Nash equilibrium. In order to find ESS's of our matching game, we propose using a novel, fast evolutionary game dynamics motivated by Darwinian selection processes, which let the pure strategies play against each other until an equilibrium is reached. A distinguishing feature of the proposed framework is that it allows one to naturally deal with general many-to-many matching problems even in the presence of asymmetric compatibilities. The potential of the proposed approach is demonstrated via two sets of image matching experiments, both of which show that our results outperform those obtained using well-known domain-specific algorithms.

3.1 Game-theoretic matching

The problem of finding correspondences within a set of elements, or features, is central to any recognition task where the object to be recognized is naturally divided into several parts. In this contexts graph-based representations have been used with considerable success do to their ability to capture concisely the relational arrangement of object primitives, in a manner which can be invariant to changes in object viewpoint.

However, applications in which estimating a set of correspondences is a central task toward the solution range from object recognition, to 3D registration, to feature tracking, to stereo reconstruction [102, 27, 93]. Several matching algorithms have been proposed in the literature. Some can just be classified as *ad hoc* solutions to specific problems, but the vast majority cast the problem into an energy minimization framework and extract approximate optimizers of an objective function within a set of feasible correspondences.

In general, the overall goal is to maximize the global or local coherence of the matched

pairs with respect to some compatibility. For example, when the problem is cast into a graph-matching approach, we can maximize the total similarity of matched nodes [21, 125], while when dealing with point-pattern matching under rigid or affine transformations we can maximize the coherence with respect to a global fitting transformation [102, 108]. Further, the globality constraint to the transformation function can be relaxed by applying it only to feature points that are close to one another, allowing for transformations that are only locally affine [114].

In most cases the objective function can be written as a monotonic transformation of the sum of pairwise interactions between matching hypotheses. This can be either the similarity between matched features, as in the graph-matching case [55, 168, 15], or due to the similarity between the underlying transformations, as for the point-pattern matching case. In the latter case the matching approach is dual to several robust parameter estimation algorithms such as RANSAC or general voting algorithms. See for example [13], where a pairwise coherence measure and a matching approach is proposed to estimate symmetries in 3D objects.

Further, quite often the set of feasible correspondences can be defined using only unary and binary relations. For instance, it is possible to guarantee a global one-to-one match and structural coherence using the association graph technique described by Barrow and Burstall [21]. Also adjacency and hierarchical constraints can be enforced on a local pairwise basis, as shown by the many techniques that cast the matching problem to an equivalent clique search in an auxiliary association graph [125, 128, 162, 160]. Formulations that satisfy these conditions range from bipartite matching, to subgraph isomorphism, to quadratic assignment, to edit-distance, and include a dual form of parameter estimation approaches such as Hough transform and RANSAC.

In the following, we present a game-theoretic approach to correspondence estimation derived from a clustering approach presented in [164]. The proposed approach is quite general since it can be applied to any formulation where both the objective function and the feasible set can be defined in terms of unary and pairwise interactions.

The main idea is to model the set of possible correspondences as a set of game strategies.

Specifically, we formulate the matching problem as a non-cooperative game where the potential associations between the items to be matched correspond to strategies, while payoffs reflect the degree of compatibility between competing hypotheses. Within this formulation, the solutions of the matching problem correspond to evolutionary stable states (ESS's), a robust population-based generalization of the notion of a Nash equilibrium. A distinguishing feature of the proposed framework is that it allows one to naturally deal with general many-to-many matching problems even in the presence of asymmetric compatibilities.

We will show that by expressing compatibilities between associations as payoffs between strategies we are able to model a class of utility functions where global maximums corresponds to stable states with maximal average payoff. Moreover, by setting a zero payoff between two associations we are able to enforce hard constraints on the match itself. Intuitively this approach allows to search for matches that not only enforces the

constraints posed, but also are made up of associations with a high level of pairwise compatibility.

In addition the percentage of population related to each strategy when in an evolutionary stable state could give an important insight about the degree of participation of each association to the match. In section 3.1.3 we suggest an optimization technique that can be used in order to find stable states in an efficient way. Finally, in section 3.2 we evaluate the effectiveness of our approach by applying it to two rather different pattern matching scenarios and by comparing the results achieved with respect to those obtained by other domain-specific techniques.

Before going into the details of the proposed framework we need to introduce some notations and definitions that will be used throughout. Let O_1 and O_2 be the two sets of features that we want to match, we define the set of *feasible associations* $\mathbb{A} \subseteq O_1 \times O_2$ the set of relations between O_1 and O_2 that satisfy the unary constraints. Hence, each feasible association represents a possible matching hypothesis. We assume that we can compute a set of *pairwise compatibilities* $C : \mathbb{A} \times \mathbb{A} \rightarrow \mathbb{R}^+$ that measure the support that one association gives to the other. Here, the self compatibilities, i.e., the compatibilities that an association gives to itself, are assumed to be zero.

In this formulation, a *submatch* (or simply a *match*) is intuitively a set of associations, which satisfies the pairwise feasibility constraints, and two additional criteria: *high internal compatibility*, i.e. the associations belonging to the match are mutually highly compatible, and *low external compatibility*, i.e. associations outside the match are scarcely compatible with those inside. This definition of match allows us to abstract from the specific problem, since domain-specific information is confined to the definition of the compatibility function. Further, we are able to deal with many-to-many, one-to-many, many-to-one and one-to-one relations in an uniform way, as we do not impose restriction on the way the associations are selected, but incorporate the constraints with the compatibilities.

3.1.1 Matching as a non-cooperative game

Following [164], we define a *matching game*. Assume that we have two sets of objects O_1 and O_2, and a compatibility function C. Two players with complete knowledge of the setup play by simultaneously selecting an association. After both have shown their choices, each player receives a payoff, monetary or otherwise, proportional to the compatibility of the selected association with respect to the association chosen by the opponent. Clearly , it is in each player's interest to pick an association, which is strongly supported by the association that the adversary is likely to choose and, assuming no prior knowledge of the inclination of the adversary, the best strategy for a player becomes the selection of associations belonging to strongly supported match.

Let $O = \{1, \ldots, n\}$ be the enumeration of the set of associations $\mathcal{A}$, where $n = |\mathcal{A}|$. Here, O is the set of *pure strategies* (in the language of game-theory) available to the players and $C = (c_{ij})$ is an $n \times n$ payoff (or utility) matrix [173], where c_{ij} is the payoff that a player gains when playing the strategy i against an opponent playing strategy j.

A *mixed strategy* is a probability distribution $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$ over the available strategies in O. Mixed strategies clearly lie in the standard simplex Δ of the n-dimensional Euclidean space, which is defined as

$$\Delta = \left\{ \mathbf{x} \in \mathbb{R}^n : \sum_{i=1}^{n} x_i = 1 \text{ and } x_i \geq 0,\ i = 1, \ldots, n \right\}.$$

The *support* of a mixed strategy $\mathbf{x} \in \Delta$, denoted by $\sigma(\mathbf{x})$, defines the set of elements with non-zero probability: $\sigma(\mathbf{x}) = \{i \in O : x_i > 0\}$. The expected payoff that a player obtains by playing the pure strategy i against an opponent playing a mixed strategy $\mathbf{x}$ is $(C\mathbf{x})_i = \sum_j c_{ij} x_j$. Hence, the expected payoff received by adopting a mixed strategy $\mathbf{y}$ is $\mathbf{y}^T C\mathbf{x}$. The *best replies* against a mixed strategy $\mathbf{x}$ is the set of mixed strategies defined as $\beta(\mathbf{x}) = \{\mathbf{y} \in \Delta : \mathbf{y}^T C\mathbf{x} = \max_{\mathbf{z}} \mathbf{z}^T C\mathbf{x}\}$. $\mathbf{x}$. It can be shown that, if $\mathbf{y}$ is in $\beta(\mathbf{x})$, then each strategy in $\sigma(\mathbf{y})$ is in $\Omega(\mathbf{x})$. A mixed strategy $\mathbf{x}$ is a *Nash equilibrium* if it is a best reply to itself, i.e. $\forall \mathbf{y} \in \Delta$, $\mathbf{y}^T C\mathbf{x} \leq \mathbf{x}^T C\mathbf{x}$. This implies that for all $i \in \sigma(\mathbf{x})$, $(C\mathbf{x})_i = \mathbf{x}^T C\mathbf{x}$, hence the payoff of every strategy in the support of $\mathbf{x}$ is constant , while all strategies outside the support of $\mathbf{x}$ earn a payoff that is less than or equal $\mathbf{x}^T C\mathbf{x}$.

Within our matching setting, Nash equilibria are good candidates for a match, as they satisfy both the internal and external compatibility criteria. In fact, any association $i \in \sigma(\mathbf{x})$ of a Nash equilibrium $\mathbf{x}$ receives from $\mathbf{x}$ the same expected payoff $(C\mathbf{x})_i = \mathbf{x}^T C\mathbf{x}$, while associations not in $\sigma(\mathbf{x})$ receive a lower or equal support from associations of the match. Note, however, that external criteria are not strict: there could exist associations not in $\sigma(\mathbf{x})$ that earn a payoff equal to $\mathbf{x}^T C\mathbf{x}$ like associations in the group, which may lead to a non isolated Nash equilibrium and, thus, an ambiguous match. Therefore, here we undertake an evolutionary game-theoretic analysis of the possible strategies available to each player.

Evolutionary game theory considers an idealized scenario wherein pairs of individuals are repeatedly drawn from a large population to play a two-player symmetric game. Each player is not supposed to behave rationally or have a complete knowledge of the details of the game, but he acts according to a pre-programmed pure strategy and a selection process allows "fit" individuals (i.e., those selecting strategies with high support) to thrive, while driving "unfit" ones to extinction. In our setup, we expect the individuals pre-programmed to select associations within a match to survive the selective pressure.

A strategy $\mathbf{x}$ is said to be an *Evolutionary Stable Strategy* (ESS) if it is a Nash equilibrium and for each best reply $\mathbf{y}$ to $\mathbf{x}$, i.e. such that $\mathbf{y}^T C\mathbf{x} = \mathbf{x}^T C\mathbf{x}$, we have $\mathbf{x}^T C\mathbf{y} > \mathbf{y}^T C\mathbf{y}$. Intuitively, ESS's are strategies such that any small deviation from them will lead to an inferior payoff.

3.1.2 Enforcing hard constraints

A main characteristic of the proposed approach is that associations pairs that have zero compatibility cannot be in the same selected submatch. this means that pairwise con-

straints can be enforced by forcing to zero the compatibility between associations that do not satisfy the constraints.

Theorem 1. *Consider a matching-game with compatibilities $C = (c_{ij})$. If $\mathbf{x} \in \Delta$ is an ESS then $c_{ij} > 0$ for all $i, j \in \sigma(\mathbf{x})$.*

Proof. Assume $c_{ij} \leq 0$ for distinct $i, j \in \sigma(\mathbf{x})$, and let $\mathbf{y} = \delta(\mathbf{e}_i - \mathbf{e}_j) + \mathbf{x}$, where $0 < \delta \leq x_j$ and $\mathbf{e}_k$ is a vector with entry k equal to one and all other entries equal to zero. Note that $\mathbf{y}$ is a best reply to $\mathbf{x}$, in fact

$$\mathbf{y}^T C\mathbf{x} = \delta(\mathbf{e}_i - \mathbf{e}_j)^T C\mathbf{x} + \mathbf{x}^T C\mathbf{x} = \mathbf{x}^T C\mathbf{x}\,,$$

where $(\mathbf{e}_i - \mathbf{e}_j)^T C\mathbf{x} = 0$ by the Nash condition on $\mathbf{x}$. However,

$$\begin{aligned}(\mathbf{x} - \mathbf{y})^T C\mathbf{y} &= -\delta(\mathbf{e}_i - \mathbf{e}_j)^T C\left[\mathbf{x} + \delta(\mathbf{e}_i - \mathbf{e}_j)\right] \\ &= -\delta^2(\mathbf{e}_i - \mathbf{e}_j)^T C(\mathbf{e}_i - \mathbf{e}_j) \\ &= -\delta^2(c_{ii} + c_{jj} - 2c_{ij}) = 2\delta^2 c_{ij} \leq 0\,,\end{aligned}$$

which contradicts the evolutionary stability of $\mathbf{x}$. □

Theorem 1 shows that if we set a non positive compatibility between two associations, then there exists no match containing them. This provides a way for expressing hard constraints in our matching framework. For instance this ability can be used to avoid one-to-many, many-to-one and many-to-many correspondences in the final match. Further, it can be exploited to model problems where other kind of pairwise constraints have to be enforced. One classical example could be the Stable Marriage Problem [75, 74], that we can tackle by assigning zero payoff to the pairs of marriages that break the stability condition. This way we are guaranteed that each ESS corresponds to a stable marriage. Moreover, it is easy to see that if we assign to each pair of feasible strategies a payoff of one, all the complete stable matches (that are guaranteed to exist [75]) correspond to ESS's that exhibit the same maximum average payoff. Finally, by choosing different payoff functions, we can model other instance of the problem.

3.1.3 Computing ESS's

In order to extract ESS's we make use of a new population game dynamics [32, 33], which is motivated by the analogy with infection and immunization processes within a population of "players". The selection mechanism governing this dynamics iteratively performs an infection step, which consists of spreading (or suppressing) the most successful (unsuccessful) strategies in the population. The infection phase is then protracted as long as the selected "infective" strategy performs better (or worse, if not extinct) than the average population's payoff. Let $\tau_+ = \{i \in O \,:\, (C\mathbf{x})_i > \mathbf{x}^T C\mathbf{x}\}$, $\tau_- = \{i \in O \,:\, (C\mathbf{x})_i < \mathbf{x}^T C\mathbf{x}\}$ and

$$\begin{aligned}\mathcal{M}(\mathbf{x}) \in \arg\max \big\{&(C\mathbf{x})_i - \mathbf{x}^T C\mathbf{x} \,:\, i \in \tau_+(\mathbf{x})\big\} \cup \\ &\big\{-(C\mathbf{x})_i + \mathbf{x}^T C\mathbf{x} \,:\, i \in \tau_-(\mathbf{x}) \cap \sigma(\mathbf{x})\big\}\,.\end{aligned}$$

Then, the dynamics are governed by

$$\mathbf{x}^{(t+1)} = \widetilde{\delta}_{\mathcal{S}(\mathbf{x}^{(t)})}(\mathbf{x}^{(t)})[\mathcal{S}(\mathbf{x}^{(t)}) - \mathbf{x}^{(t)}] + \mathbf{x}^{(t)}\,, \tag{3.1}$$

where

$$\mathcal{S}(\mathbf{x}) = \begin{cases} \mathbf{e}_i & i = \mathcal{M}(\mathbf{x}) \in \tau_+(\mathbf{x}x) \\ \frac{x_i}{x_i - 1}\,(\mathbf{e}_i - \mathbf{x}) + \mathbf{x} & i = \mathcal{M}(\mathbf{x}) \in \tau_-(\mathbf{x}) \cap \sigma(\mathbf{x}) \\ \mathbf{x} & \text{otherwise} \end{cases}$$

and

$$\widetilde{\delta}_{\mathbf{y}}(\mathbf{x}) = \begin{cases} \min\left[1, \frac{(\mathbf{x}-\mathbf{y})^T A \mathbf{x}}{(\mathbf{y}-\mathbf{x})^T A (\mathbf{y}-\mathbf{x})}\right] & (\mathbf{y}-\mathbf{x})^T A (\mathbf{y}-\mathbf{x}) < 0 \\ 1 & \text{otherwise.} \end{cases}$$

This evolution process exhibits a number of nice properties: first the asymptotically stable points for the dynamics are the ESS's. Second, it is computationally very efficient, as each iteration has linear time complexity.

3.2 Experimental results

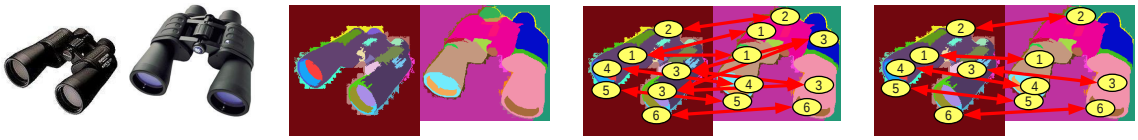
In order to evaluate both the generality and the effectiveness of the proposed approach we performed two sets of experiments. In the first set of experiments we match the segmentations of images with similar subjects. In this context the intrinsic instability of the extracted segments requires a more relaxed coherence constraint, moreover the presence of under- and over-segmentation requires the matches to be many-to-many. In the second set we use our game-theoretic framework to match point-patterns extracted from images after affine transformation. Here we expect the features to be stable and preserved after the transformation. However, the approach must be robust against the appearance of outliers. For this particular application we are requiring strict one-to-one correspondences and a tight global coherence between matched points, but given the presence of a clear ground truth, a more quantitative analysis is performed. All the experiments where run on a standard PC with a 2GHz processor.

3.2.1 Segmentation Matching

The first set of experiments assesses the effectiveness of the approach on an object recognition task. Here, we match similar objects from the Caltech-256 database [71], which exhibits large variations in illumination, scale and viewpoint. In this context invariant feature points cannot be exploited as they are not robust with respect to changes in the appearance of objects belonging to the same semantic category. Thus we employed more robust, but less repeatable features: We segmented the images using the algorithm presented in [58] and matched the corresponding segments, in a process similar to [78].

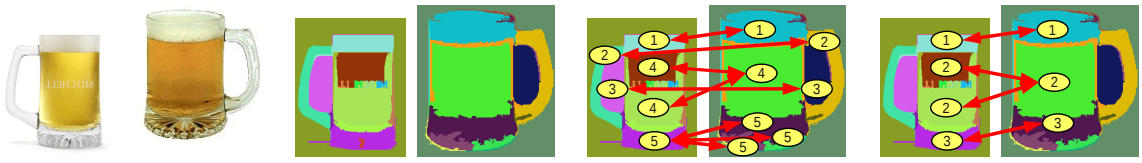
Starting from relatively unstable segmentations, we selected candidate matches computing the normalized cross-correlation of each segment pair, and we selected the top 10

$\alpha = 0.05\ \pi = 0.57$ $\alpha = 0.05\ \pi = 0.44$



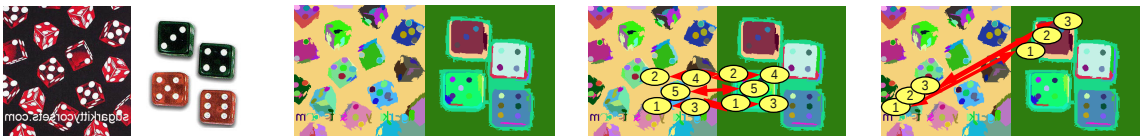
$\alpha = 0.05\ \pi = 0.54$ $\alpha = 0.05\ \pi = 0.51$

Many-to-many vs One-to-one

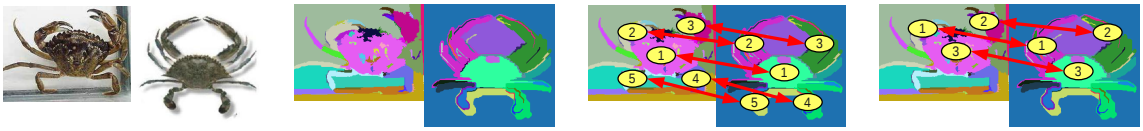


$\alpha = 0.05\ \pi = 0.57$ $\alpha = 0.05\ \pi = 0.49$

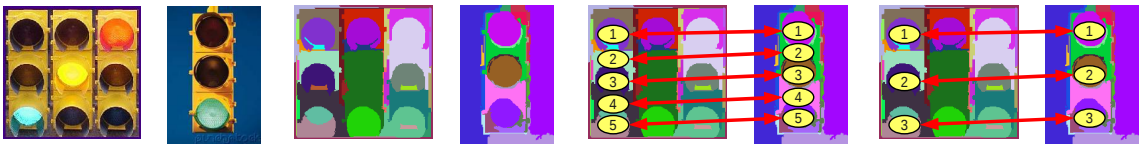
Chirality constraint



$\alpha = 0.1\ \pi = 0.58$ $\alpha = 0.5\ \pi = 0.55$



$\alpha = 0.1\ \pi = 0.41$ $\alpha = 0.5\ \pi = 0.21$



$\alpha = 0.1\ \pi = 0.45$ $\alpha = 0.5\ \pi = 0.24$

Effect of selectivity parameter α

Figure 3.1: Region matching: the first four columns show the original images and the extracted segments, while columns five to eight show the resulting matches. The first two rows show the result of enforcing one-to-one correspondence, the third row show the result of enforcing chirality (handedness) of the matching segments, while the last three rows show the effect of changes in the selectivity parameter α.

local maxima as possible matching candidates. In order to account for chiral segments, we also computed the normalized cross-correlation against the mirror image of each segment. To measure the support that each match obtains from other matches we computed the payoff matrix $C = (c_{ij})$ as follows: Given the normalized cross correlation values v_i and v_j at the local maxima associated to candidate matches i and j respectively, and the displacement vectors $\mathbf{t_i}, \mathbf{t_i}$ between each template and the corresponding matching segment, we defined the coherence between candidates as $c_{ij} = v_i v_j e^{\alpha|\mathbf{t_i}-\mathbf{t_j}|}$. Here α is a selectivity parameter that affects the decay of the coherence and thus the selectivity of the match to be found.

In addition to the continuous parameter α we also enforced two hard constraints: namely the one-to-one relationship in the matches and the chirality constraint, which forces the matches to have the same handedness. The first constraint is obtained by setting to 0 the coherence of any pair of candidates where either the source segment or the destination segment coincide, while the second is similarly enforced by setting to 0 the coherence of pair of candidates in which one segment maps to a mirrored segment while the other maps to a straight one.

Figure 3.1 shows the results on a few selected shapes for which the categorization performance presented in [71] was around the middle in the rank order. For each row the first two columns show the test images, while the third and fourth column show the extracted segments, while the fifth, sixth, seventh and eight columns show the matches obtained together with the values of the parameter α used and the resulting average payoff pi.

The first two rows show the effect of enforcing the one-to-one constraint versus allowing a full many-to-many match. Here the fifth and sixth column show the results with a full many-to-many match, while the seventh and eight column show the effect of enforcing the one-to-one match. Note how in the first row the second beer is over-segmented due to a writing on the glass. By enforcing a one-to-one correspondence only part of the region is selected, while allowing many-to-many matches all the segments are mapped to the equivalent segment on the other image. Further, in the binoculars example, the part on the right is segmented differently on the two images and all the corresponding segments on the first image are mapped to the segments on the second thus giving a full many-to-many match. On the other hand, enforcing a one-to-one correspondence we are not able to match the top part of the optics.

The third row shows the effect of the chirality constraint. Here the fifth and sixth columns show the best match obtained using the mirrored segments, while the seventh and eight columns show the result obtained eliminating the mirror candidates. Note that in the latter case only the symmetric part of the shape is matched.

The last three rows show the effect of increasing the selectivity parameter α. Note that increasing alpha forces the approach to select matches that are more geometrically coherent, even when this results in fewer segments matched and a lower average payoff.

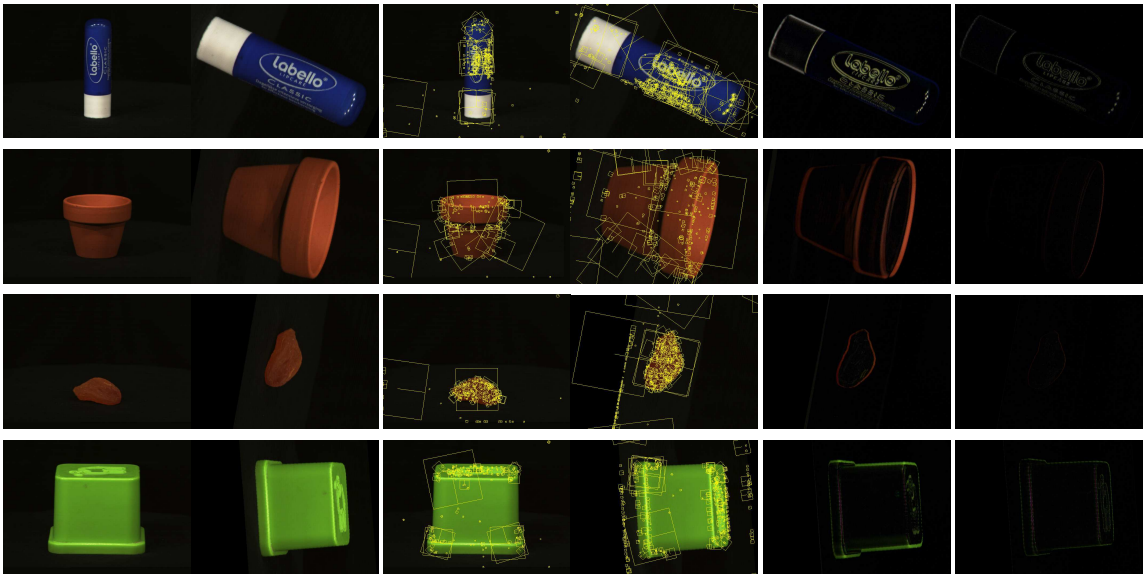


Figure 3.2: Point pattern matching: the first two columns show the original images, the third and fourth columns show the extracted features, and the fourth and fifth show the alignment error using the transforms estimated using ransac (fifth) and our approach (sixth).

3.2.2 Point-Pattern Matching

In this set of experiments our goal is to test the ability of the proposed framework to match corresponding features points between two instances of the same image with modified scale and orientation. The features points are extracted from each image with the widely used SIFT algorithm [102]. SIFT features are known to be highly repeatable under a large class of affine transformations and are very resilient to splitting or joining. Under these conditions we need a very selective matcher which enforces a common global transformation to all the matched features. In [102] Lowe gauges the coherence of the transformation using ransac. This, however, requires a global threshold for the consensus, which limits the precision of the estimation.

The experiments where performed on the Aloi database [62]. For each run we selected 20 images and randomly deformed them with an affine transformation with a scale variation between 0.5 and 2 and a rotation between 0.5 and 2.0 radians. We extracted the SIFT features from the original and transformed image and picked as candidate associations all the pairs with sufficiently similar descriptors.

Each candidate association represents a single transformation and supports only associations with similar transformations. To measure the support between two associations, we project the first point of one association with the transformation of the other association. Then we measure the distance between the transformed point and the corresponding point in the first association. We repeat the operation reversing the role of the two associations obtaining the two distances d_1 and d_2. The support is, then, $e^{-\max(d_1,d_2)}$. Once

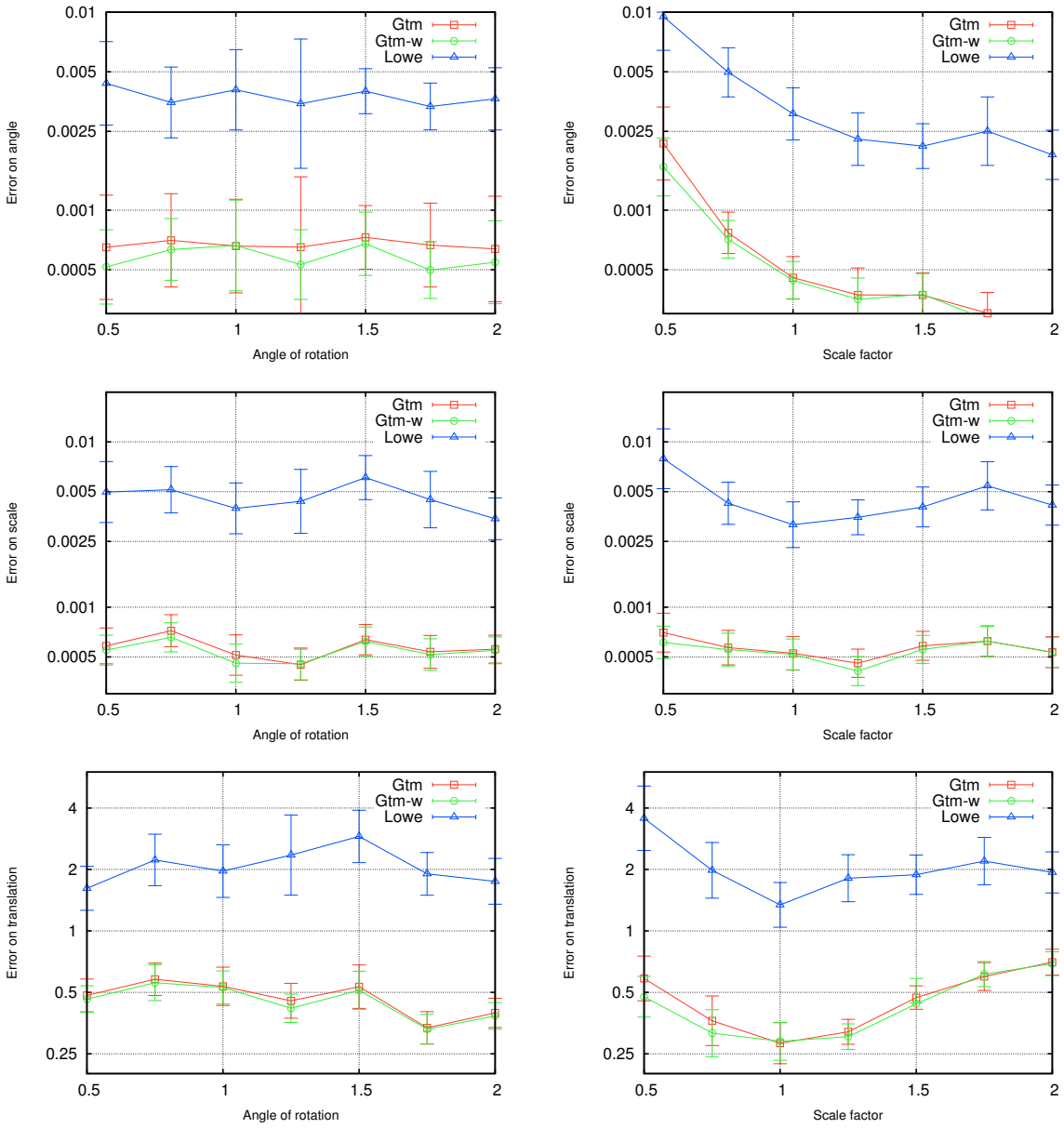


Figure 3.3: Point pattern matching: Error in the estimation of translation, scale and rotation as we increase the variations in scale and orientation.

the best match is extracted, we have two alternatives to compute the final transformation: the first is an unweighted approach where we compute a simple average of the transformation parameters related to the associations in the match. The second approach weighs the transformation parameters with the proportion of the population playing the related strategy at equilibrium.

We compare our approach with ransac, where we determine the associations to agree within tolerance if $\max(d_1, d_2) < 5$ pixels. The value of 5 pixels was experimentally

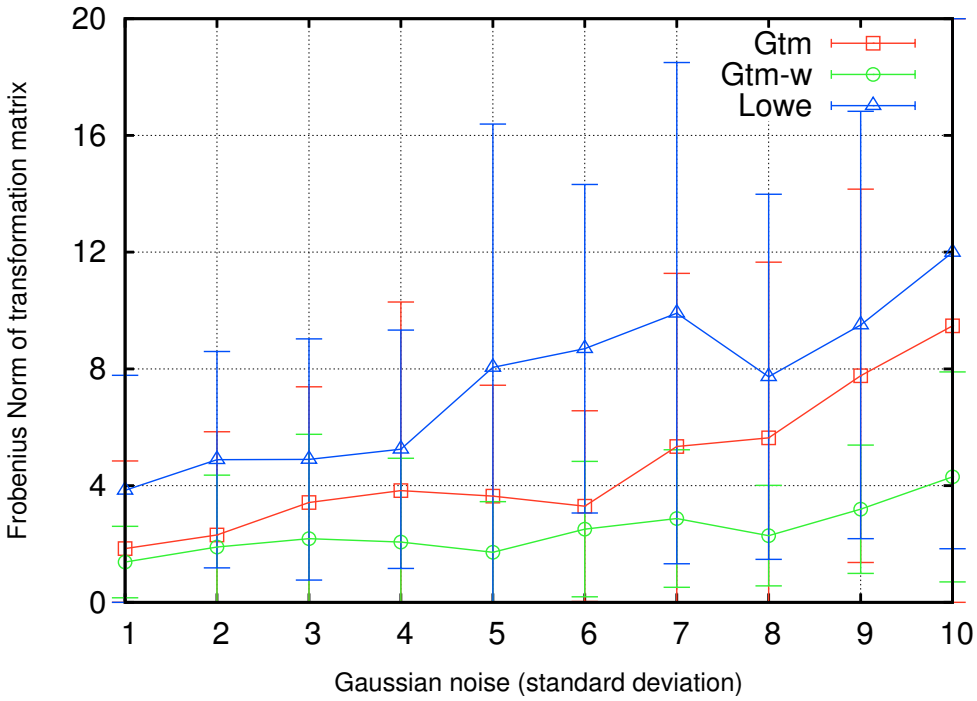


Figure 3.4: Sensitivity to noise: Frobenius norm of the difference between exact and estimated transformation errors under an increasing amount of Gaussian noise.

determined to be the one which gave the best results. Note that this threshold on the error limits the accuracy of ransac, while our approach, being parameter-less, does not suffer from this drawback.

Figure 3.2 shows the original images (first two columns), the extracted features (third and fourth column), and the transformation error obtained using the two approaches (last two columns). The error is the difference between the original image transformed with the estimated transformation and the second image. The fifth column shows the error obtained using the transformation estimated with ransac, while the sixth column shows the difference using the transformation estimated using the weighted version of our approach. As it can be seen our approach estimates the transformation with higher accuracy than ransac. So much so that the difference images are almost completely black. This is mainly due to the lack of a lower bound on the precision of the transformation, which for ransac is enforced by the consensus threshold.

Figure 3.3 plots the error in the estimation of translation, scale and rotation as we increase the variations in scale and orientation. The average and standard deviations are computed over 140 images. As it can be seen, the weighted and unweighted versions of our approach have similar performance, with the weighted version exhibiting slightly lower error. On the other hand ransac show errors an order of magnitude larger in all conditions.

In an attempt to quantify the sensitivity of the approach to noise, we added an increas-

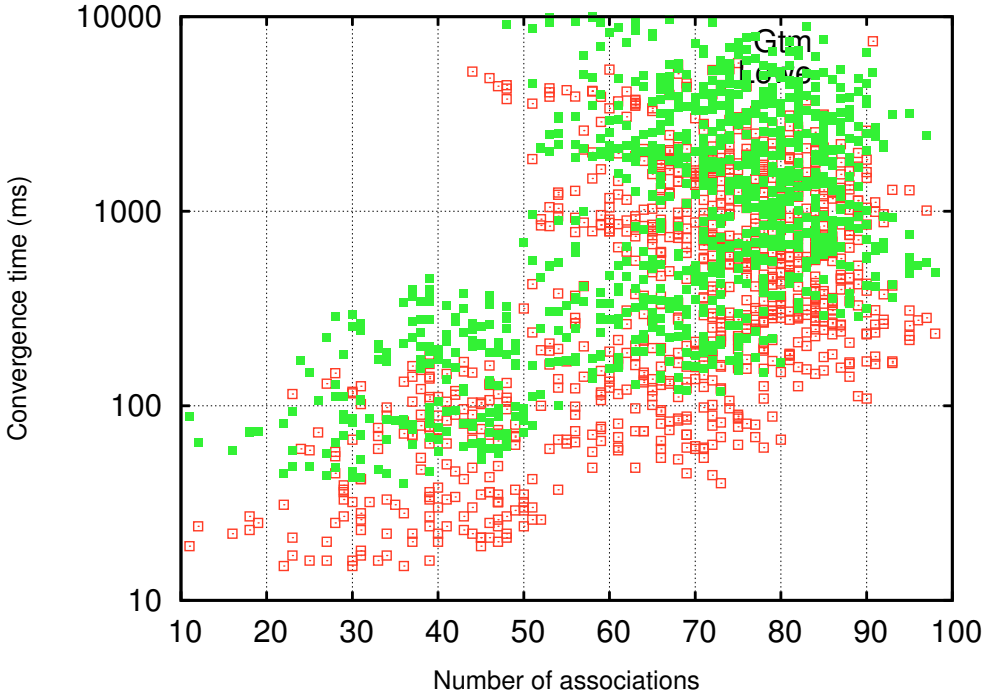


Figure 3.5: Scatter plot of feature points versus computation time (msecs).

ing amount of Gaussian noise to the rotated and scaled images before we computed the SIFT features. This introduces an increasing number of outliers as well as missing feature points. Figure 3.5 plots the Frobenius norm of the difference between the ground truth and the estimated transformation matrices as the standard deviation of the Gaussian noise increases. For each noise level we selected 20 images and randomly deformed them with an affine transformation with a scale variation between 0.5 and 2 and a rotation between 0.5 and 2.0 radians. From the plot we can see that our approach maintains a much lower error as compared to ransac even at high noise levels. Further, we can see that, while the rate with which the error increases with noise is similar for ransac and the unweighted version of our approach, the weighted version appears to provide much lower error even with a high level of noise.

Figure 3.5 shows a scatter plot of number of feature-points versus runtime for our approach (green) versus ransac (red). As it can be seen, ransac is slightly faster, with a geometric average over all runs of 314 msecs. While our approach has a geometric average of 762 msecs, both method having a relative deviation (over all the experiments) of approximately 450%. Further, the scatter plot confirms the finding of a factor 2.4 slowdown with our approach, arguably providing a favorable accuracy/performance ratio.

3.3 Concluding Remarks

In this chapter we presented a game-theoretic approach for the all-pervasive matching problem in computer vision when both the objective of the match and the feasible set can be defined based on pairwise interactions. We formulated the matching problem as a non-cooperative game between matching hypotheses, while payoffs reflect the degree of compatibility between associations. Within this formulation, the solutions of the matching problem correspond to evolutionary stable states (ESS's), a robust population-based generalization of the notion of a Nash equilibrium. A distinguishing feature of the proposed framework is that it allows one to deal uniformly with many-to-many, many-to-one and one-to-one matching approaches as well as robust estimation of a parametrized matching transformation. The potential of the proposed approach has demonstrated via two sets of image matching experiments, both of which show that our approach outperforms well-known algorithms at the state of the art.

4

Modeling Higher-Order Problems

In this chapter we present three adaptation of our framework to situations where it becomes necessary to deal with compatibilities defined between more than two items and to preserve some structural relationship between matched pairs. Specifically we will show how to match attributed graphs by using an attributed association graph and searching for a maximum weight clique in it in section 4.1. In section 4.2 we will track articulated objects by defining a special association graph that enforces structure preservation. Finally, in section 4.3, we will offer an alternative view of higher order matching by modeling compatibilities between sets of matches as hyper-edges in a hyper-graph.

4.1 Encoding Higher-Orderness with Weighted Graphs

The matching of relational structures is a problem that pervades computer vision and pattern recognition research. A classic approach is to reduce the matching problem into one of search of a maximum clique in an auxiliary structure: the association graph. The approach has been extended to incorporate vertex-attributes by reducing it to a weighted clique problem, but the extension to edge-attributed graphs has proven elusive. However, in vision problems, quite often the most relevant information is carried by edges. For example, when the graph abstracts scene layout, the edges can represent the relative position of the detected features, which abstracts the geometry of the scene in a way that is invariant to rotations and translations. In this section, we provide a generalization of the association graph framework capable of dealing with attributes on both vertices and edges. Experiments are presented which demonstrate the effectiveness of the proposed approach.

4.1.1 Matching with graphs

Graph-based representations have long been used with considerable success in computer vision and pattern recognition in the abstraction and recognition of objects and scene structure. Concrete examples include the use of shock graphs to represent shape-skeletons [94, 145, 141], the use of trees to represent articulated objects [83, 187] and the use of Delaunay graph to represent the distribution of features in a scene [175]. The

attractive feature of structural representations is that they concisely capture the relational arrangement of object primitives, in a manner which can be invariant to changes in object viewpoint. Using this framework we can transform a recognition problem into a relational matching problem.

The problem of how to measure the similarity or distance of pictorial information using graph abstractions has been a widely researched topic of over twenty years. Early work on the topic includes Barrow and Burstall's idea [21] of locating matches by searching for maximum common subgraphs and Shapiro and Haralick's idea [143] of locating the isomorphism that minimizes the weight of unmapped nodes. A common approach transforms the combinatorial problem into a continuous optimization problem and then uses the wide range of available optimization algorithms available to find an approximate solution. In [43] Kittler, Christmas, and Petrou use relaxation labelling to label nodes in the data graph with the corresponding node in the model graph and use graph connectivity to combine evidence. Relaxation labelling, however, does not guarantee a one-to-one correspondence between nodes. In order to guarantee a one-to-one assignment, Gold and Ragaranjan [66] introduced the "graduated assignment" method. This is an evidence combining model that guarantees two way constraints. Evidence combining methods like relaxation labelling and graduated assignment give a very interesting framework to iteratively improve on our initial estimate, but they are critically dependent on a good consistency model and a reliable initialization.

Another classic approach pioneered by Ambler et al. [16] is to reduce the matching problem into one of search of a maximum clique in an auxiliary structure: the association graph. This graph is defined over a vertex-set that is the Cartesian product of the vertex-sets of the original structures, and the edges represent the compatibility of two maps between the original graphs. Optimal matches between the two structures are then in a one-to-one relationship with maximum cliques on the association graph. By re-casting the search for the maximum common subgraph as a max clique problem [21], we can tap into a diverse array of powerful heuristics and theoretical results available for solving the max clique problem. An important development in that direction is reported by Pelillo [125] who, using the Motzkin-Straus theorem [54] transforms the max clique problem into a continuous quadratic programming problem, and shows how relaxation labelling can be used to find an approximate solution.

By adding a weight to the vertices of the association graph, the similarity between vertices can be taken into account. This allows us to match vertex-attributed graphs by searching for cliques of maximum weight [127]. However, in vision problems, quite often the most relevant information is carried by edges. For example, when the graph abstracts scene layout, the edges can represent the relative position of the detected features, which abstracts the geometry of the scene in a way that is invariant to rotations and translations. Furthermore, any deformation of the scene will induce changes in the relative distances of the features, leaving the actual features mostly unchanged. Examples of such representations include the use of Delaunay graphs for scene registration [175] and the use of skeletons for shape recognition [141]. In these structural representations it is the edge that takes center stage, and, hence, it is essential for a matching algorithm to deal with

edge and vertex attributes in a uniform and well founded way. To this end we propose to extend the association graph framework to deal with edge associations directly, without having to infer them from vertex associations. This way edge similarity information can be incorporated into the matching process.

A drawback of this approach is that the size of the association graph increases with the product of the number of edges in the two graphs, making the space and time requirements of the Motzkin-Straus formulation too demanding. For this reason we opted for the use of Reactive local search (RLS) [23], a search-based clique heuristic which can find a candidate solution quickly and then refine it as requested. This characteristic, which is common to most search-based heuristics, makes the approach usable for range queries and any-time queries in a structural database, i.e, queries where only an upper bound on the actual distance is required and interactive queries that can be stopped as soon as the user is satisfied with the results. However, the size requirement still limits us to sparse graphs. Fortunately, most scene abstractions use planar graph and, hence, are sparse.

4.1.2 Edge Association Graph

Let $G = (V, E)$ be an undirected graph with vertex set V and edge set $E \subset V^2$, any time there is an arc between nodes $u, v \in V$ we say that the nodes are *adjacent* and we write $u \sim v$. If $G = (V, E)$ is a directed graph, with the notation $u \rightsquigarrow v$ we indicate that there is an edge from node $u \in V$ to node $v \in V$; furthermore, we call a *self loop* an arc from one node to itself. Let $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ be two directed graphs, a subgraph isomorphism f between G_1 and G_2 is a partial injective function from V_1 to V_2 that respects the adjacency of the two graph, i.e.,

$$\forall u, v \in \mathrm{dom}(f),\ u \rightsquigarrow v \Leftrightarrow f(u) \rightsquigarrow f(v)\,,$$

where $\mathrm{dom}(f)$ is the subset of V_1 where f is defined.

Let $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ be two directed graphs without self loops, we define the set of correspondences between edges $V_e = E_1 \times E_2$. As each vertex can be bijectively mapped to a self loop, we can extend V_e to include vertex-correspondences by defining a generalized edge association set $V_a = V_e \cup \hat{V_1} \times \hat{V_2}$, where $\hat{V} = \{(v, v) | v \in V\}$ is the space of self loops. This way we guarantee uniformity in notation, since every entity, be it an edge or a vertex, is represented by a (possibly equal) pair of vertices.

Any subset $S \subseteq V_a$ represent a relation between edges and vertices in G_1 and edges and vertices in G_2. We define a map $\phi : \mathcal{P}(V_a) \rightarrow \mathcal{P}(V_1 \times V_2)$ from edge-relations to vertex-relation as follows:

$$\begin{aligned}\phi(X) = \{&(v_1, v_2) \in V_1 \times V_2 \,| \\ &\exists u_1 \in V_1, u_2 \in V_2\,, ((v_1, u_1), (v_2, u_2)) \in X \vee ((u_1, v_1), (u_2, v_2)) \in X\}\end{aligned} \quad (4.1)$$

ϕ is not invertible, as it is not injective, but it has a right partial inverse $\phi^{-1} : \mathcal{P}(V_1 \times V_2) \rightarrow \mathcal{P}(V_a)$ defined as follows:

$$\phi^{-1}(Y) = \{((u_1, v_1), (u_2, v_2)) \in V_a \,| (u_1, u_2) \in Y \wedge (v_1, v_2) \in Y\} \quad (4.2)$$

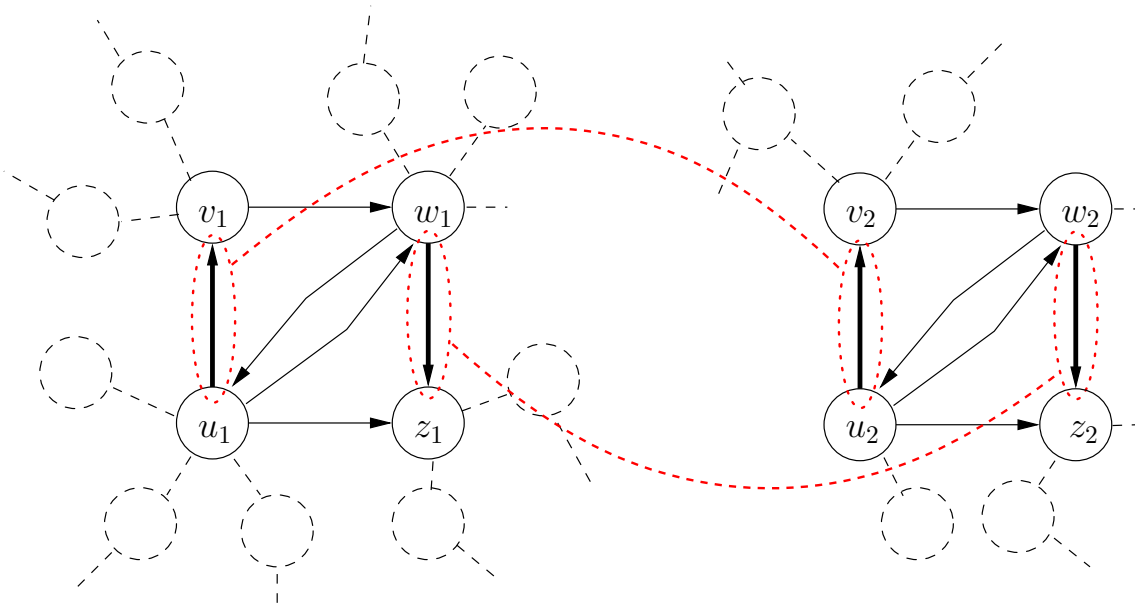


Figure 4.1: Two edge associations are compatible if they induce a subgraph isomorphism.

It is easy to show that $Y = \phi(\phi^{-1}(Y))$ and $X \subseteq \phi^{-1}(\phi(X))$. The last condition implies that for each vertex-relation there is a maximal edge-association that contains all the edge-associations inducing the same vertex-relations. Our goal is to find a set of edge associations that induce through ϕ a subgraph isomorphism between G_1 and G_2.

Let $ea1 = ((u_1, v_1), (u_2, v_2)) \in V_a$ and $ea2 = ((w_1, z_1), (w_2, z_2)) \in V_a$ be two edge-associations, $ea1$ and $ea2$ are said to be *compatible* if and only if $\phi(\{ea1, ea2\})$ is a subgraph isomorphism between G_1 and G_2, that is if the relation mapping u_1 to u_2, v_1 to v_2, w_1 to w_2, and z_1 to z_2 is a partial injective function that respects the adjacency conditions between the subgraphs of G_1 and G_2 obtained restricting them to the vertices $\{u_1, v_1, w_1, z_1\}$ and $\{u_2, v_2, w_2, z_2\}$ respectively. More formally, the edge-associations $((u_1, v_1), (u_2, v_2)) \in V_a$ and $((w_1, z_1), (w_2, z_2)) \in V_a$ are compatible if they satisfy the following relations:

$$
\begin{aligned}
u_1 = w_1 \Leftrightarrow u_2 = w_2\,, &\quad \text{(map1)}\\
v_1 = z_1 \Leftrightarrow v_2 = z_2\,, &\quad \text{(map2)}\\
u_1 = z_1 \Leftrightarrow u_2 = z_2\,, &\quad \text{(map3)}\\
v_1 = w_1 \Leftrightarrow v_2 = w_2\,; &\quad \text{(map4)}
\end{aligned}
$$

and

$$
\begin{aligned}
u_1 \rightsquigarrow w_1 \Leftrightarrow u_2 \rightsquigarrow w_2 \wedge w_1 \rightsquigarrow u_1 \Leftrightarrow w_2 \rightsquigarrow u_2\,, &\quad \text{(iso1)}\\
v_1 \rightsquigarrow w_1 \Leftrightarrow v_2 \rightsquigarrow w_2 \wedge w_1 \rightsquigarrow v_1 \Leftrightarrow w_2 \rightsquigarrow v_2\,, &\quad \text{(iso2)}\\
u_1 \rightsquigarrow z_1 \Leftrightarrow u_2 \rightsquigarrow z_2 \wedge z_1 \rightsquigarrow u_1 \Leftrightarrow z_2 \rightsquigarrow u_2\,, &\quad \text{(iso3)}\\
v_1 \rightsquigarrow z_1 \Leftrightarrow v_2 \rightsquigarrow z_2 \wedge z_1 \rightsquigarrow v_1 \Leftrightarrow z_2 \rightsquigarrow v_2\,, &\quad \text{(iso4)}\\
v_1 \rightsquigarrow u_1 \Leftrightarrow v_2 \rightsquigarrow u_2\,, &\quad \text{(iso5)}\\
z_1 \rightsquigarrow w_1 \Leftrightarrow z_2 \rightsquigarrow w_2\,. &\quad \text{(iso6)}
\end{aligned}
$$

The first four relations guarantee that ϕ induces a partial injective map, while the last six guarantee that the map induces a subgraph isomorphism. Note that the existence of the edge-associations already guarantees that $u_1 \rightsquigarrow v_1$, $u_2 \rightsquigarrow v_2$, $w_1 \rightsquigarrow z_1$, and $w_2 \rightsquigarrow z_2$, hence the lack of symmetry in the last two isomorphism conditions. Furthermore, the missing map conditions $u_1 = v_1 \Leftrightarrow u_2 = v_2$ and $u_1 = v_1 \Leftrightarrow u_2 = v_2$ are enforced by the way V_a is constructed.

Given the notion of compatibility between edge association, we define the *edge-association graph* as the undirected graph $G_a = (V_a, E_a)$ where two edge associations are adjacent if and only if they are compatible.

Proposition 1. *Let $X \subseteq V_a$, then X is a clique of G_a if and only if $\phi(X)$ is a subgraph isomorphism between G_1 and G_2.*

Sketch of proof. Let $S \subseteq V_a$ such that S is not a clique, then there must be two edge-associations $ea1, ea2 \in S$ such that $\phi(\{ea1, ea2\})$ is not a subgraph isomorphism. This implies that $\phi(S)$ is not a subgraph isomorphism and hence, by contrappositive, that the fact that $\phi(X)$ is a subgraph isomorphism implies that X is a clique of G_a. Conversely, Let $S \subseteq V_a$ such that $\phi(S)$ is not a subgraph isomorphism, then there must be two distinct vertex relations (u_1, w_2) and (v_1, w_2) that either prevent the relation from being a partial injective function, i.e., $u_1 = w_1$ or $u_2 = w_2$, or that do not respect the adjacency relations. Since $\phi(S)$ is induced by S, the two vertex relation must be induced by some edge relations, i.e., there must exist $v_1, z_1 \in V_1$ and $v_2, z_2 \in V_2$ such that either $((u_1, v_1), (u_2, v_2)) \in S$ or $((v_1, u_1), (v_2, u_2)) \in S$, and either $((w_1, z_1), (w_2, z_2)) \in S$ or $((z_1, w_1), (z_2, w_2)) \in S$, but it can be easily shown that any of these conditions must break one of the compatibility relations. □

Theorem 2. *If $X \subseteq V_a$ is a maximal clique of G_a, then $\phi(X)$ is a maximal subgraph isomorphism between G_1 and G_2. Conversely, If f is a maximal subgraph isomorphism between G_1 and G_2, then $\phi^{-1}(f)$ is a maximal clique of G_a.*

Proof. Assume that $\phi(X)$ is not maximal, but that the relation (u_1, u_2) can be added, then we can add to X the edge relation between the self loops (u_1, u_1) and (u_2, u_2), hence X cannot be maximal. Conversely, assume that we can add the edge-association $((u_1, v_1), (u_2, v_2))$ to $\phi^{-1}(f)$ then either (u_1, u_2) or (v_1, v_2) must not be in f or else $((u_1, v_1), (u_2, v_2))$ would be in $\phi^{-1}(f)$. Hence, $g = f \cup \{(u_1, u_2), (v_1, v_2)\}$ is a subgraph isomorphism and $f \subset g$, hence f is not maximal. □

This result provides us with a strong relation between maximal cliques in the edge-association graph and maximal subgraph isomorphism, note however that the relation is not a bijection, since there might be non-maximal cliques inducing a maximal isomorphism; however, these must be subsets of a maximal clique inducing the same isomorphism.

4.1.3 Weighted isomorphism

Let $\omega_e : E_1 \times E_2 \to \mathbb{R}_+$ be a similarity function between two edges in G_1 and G_2 and $\omega_v : V_1 \times V_2 \to \mathbb{R}_+$ a similarity function between two vertices in G_1 and G_2, we define the weight of an isomorphism f as

$$\Omega(f) = \sum_{\substack{(u_1,v_1) \in E_1 \\ u_1,v_1 \\ \text{dom}(f)}} \omega_e((u_1,v_1),(f(u_1),f(v_1))) + \sum_{u_1 \in \text{dom}(f)} \omega_v(u_1, f(u_1)) \,, \quad (4.3)$$

Similarly, we define a weight on a vertex $((u_1,v_1),(u_2,v_2)) \in V_a$ as

$$\omega((u_1,v_1),(u_2,v_2)) = \begin{cases} \omega_v(u_1,u_2) & \text{if } u_1 = v_1 \,, \\ \omega_e((u_1,v_1),(u_2,v_2)) & \text{otherwise.} \end{cases} \quad (4.4)$$

With these weights, we can define a weighted association graph $G_a = (V_a, E_a, \omega)$. The weight of a subset of vertices $X \subseteq V_a$ is:

$$\Omega(X) = \sum_{((u_1,v_1),(u_2,v_2)) \in X} \omega((u_1,v_1),(u_2,v_2)) \,. \quad (4.5)$$

We can now prove the following:

Proposition 2. *If f is a subgraph isomorphism between G_1 and G_2, then $\Omega(f) = \Omega(\phi^{-1}(f))$. Conversely, if $X \subseteq V_a$ is a maximal clique of G_a, then $\Omega(X) = \Omega(\phi(X))$.*

Proof. If f is an isomorphism, for each $(u_1,v_1) \in E_1, u_1, v_1 \in \text{dom}(f)$ we have $((u_1,v_1),(f(u_1),f(u_2))) \in \phi^{-1}(f)$, and for each $u_1 \in \text{dom}(f)$ we have $((u_1,u_1),(f(u_1),f(u_1))) \in \phi^{-1}(f)$. Hence each element in the sum in (4.3) will be present in one and only one node of $\phi^{-1}(f)$, which implies that $\Omega(f) = \Omega(\phi^{-1}(f))$. The second part derives from the fact that if X is maximal, then $X = \phi^{-1}(\phi(X))$. □

Finally, we obtain:

Theorem 3. *$X \subseteq V_a$ is a maximum weight clique of G_a if and only if $\phi(X)$ is a maximum weight isomorphism between G_1 and G_2. Furthermore, $\Omega(X) = \Omega(\phi(X))$.*

Proof. This derives directly from Theorem 1 and Proposition 2. □

This result allows us to cast the search for the maximum weight subgraph isomorphism between vertex- and edge-attributed graphs into an instance of the weighted clique problem in the edge-association graph.

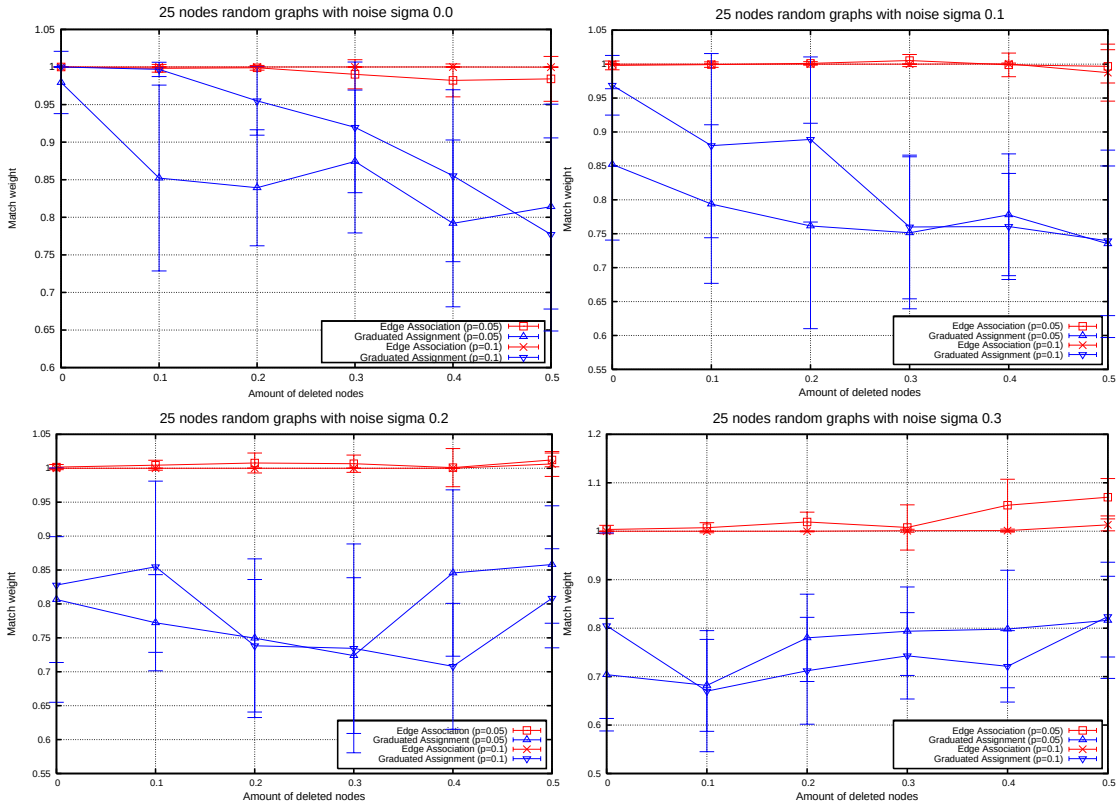


Figure 4.2: Comparison with Graduated Assignment on synthetic p-random graphs

4.1.4 Experimental Results

In order to asses the usefulness of our matching approach in recognition tasks using graph-based representations, we performed a set of experiments with synthetic as well as real-world data.

In our first set of experiments we compare matches obtained through the proposed approach with matches obtained with Graduated Assignment (GA) [66] on random graphs. For each experiment we randomly generate graph with 25 nodes and 5% and 10% densities. Vertices and edges where given real-valued attributes uniformly distributed between 0 and 1 and the similarity between two such attributes was an exponential decay of the absolute difference of the values. Each graph was perturbed by adding Gaussian noise to the attributes and by randomly removing nodes. Figure 4.2 compares the performance of the proposed approach with the results obtained by GA as both the amount of structural and of attribute noise is increased. The plot shows the ratio of the weight of the match obtained by the two algorithms over the weight of the ground-truth correspondences. Here the proposed approach clearly outperforms GA.

Next we tried to characterize the time versus performance behavior of the algorithm. To do this we generated 100 random graphs of size 25 and density 10% and ran our algorithm on them sampling at various times the best match found. To measure the overall

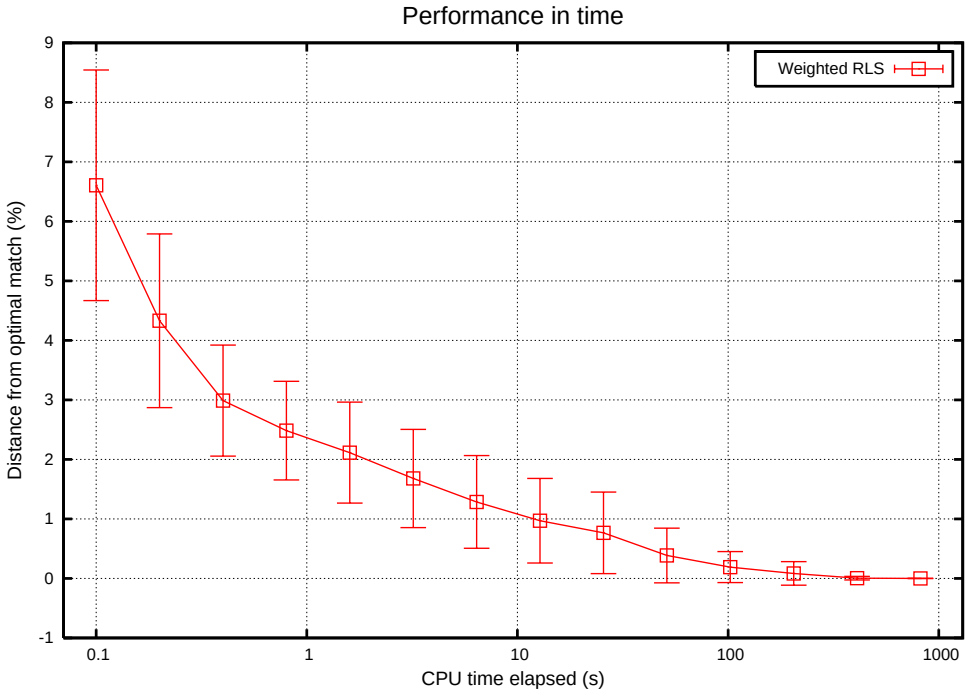


Figure 4.3: Quality of match found by early stop

quality of the match we used a metric proposed in [163]:

$$d(G_1, G_2) = 1 - \frac{\Omega(f)}{|G_1| + |G_2| - \Omega(f)}\,. \tag{4.6}$$

In Figure 4.3 we report the error of the best solution at time t relative to the best obtained letting the search algorithm run for 15 minutes. The experiments were run on PC clocked at 1.6 Ghz. From these results we see that in just 1 second the match is within 2.5% of the optimum.

We were also interested in assessing the performance of our algorithm with real-world classification problems. To this end we used a collection of 45 images from the ALOI[62] database, showing 9 objects from different viewpoints. Starting from feature-points extracted using the Harris corner detector [77], we generated attributed graphs in two different ways: The Delaunay triangulation of the points and a 5-neighbor graph, in which each vertex is connected to 5 the nearest vertices. In both cases the distances between points where used as attributes for the edges. The number of extracted vertices ranged from a minimum of 22 to a maximum of 36 and the resulting density was similar to those used in the synthetic experiments. Figures 4.5 and 4.6 show the 6 best matches for several test images as well as the corresponding value of the metric (4.6).

From the Figures we can see that the approach almost always selects similar images top matches, although, due to the instability of the representations with respect to changes in viewpoints, the algorithm was often presented with graphs with significant structural deformation.

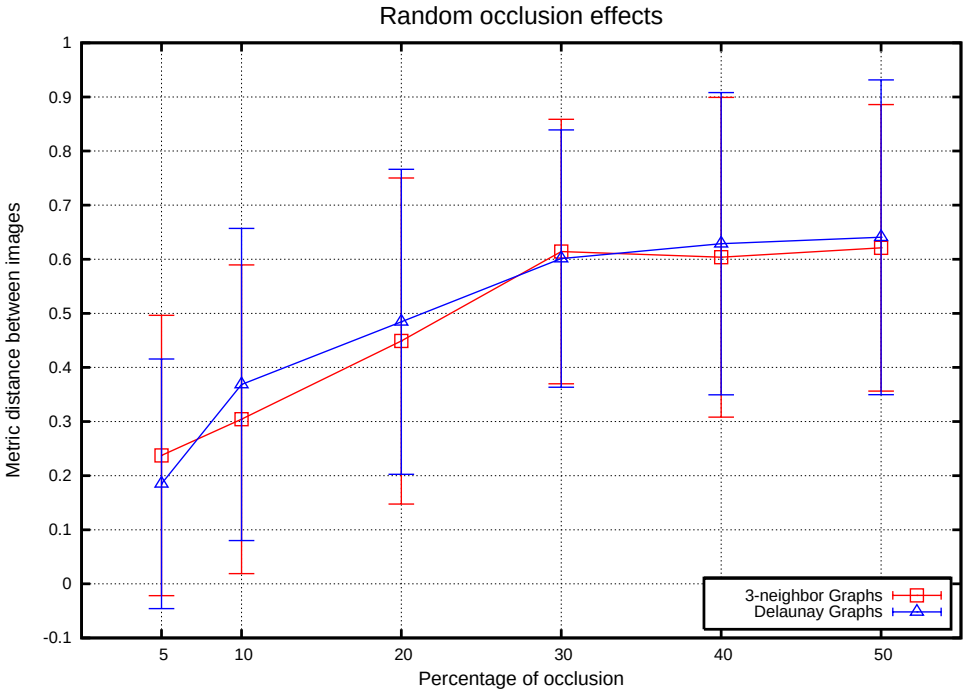


Figure 4.4: Effect of occlusions on distance

Finally we tested the robustness of our approach with respect to random occlusions on the images. To this aim we generated random occlusions ranging from 5% to 50% of the total area and ran our algorithm matching the original the graphs generated from the original image with those obtained from the occluded images. The results of this set of experiments is summarized in Figure 4.4. Here we can see that the distances from the original graph increases smoothly with the amount of occlusion, assigning reasonably small distances to images with small to moderate occlusion.

4.1.5 Concluding Remarks

In this section we presented a extension of the association graph framework which deals with edge associations directly, without having to infer them from vertex associations. This way edge similarity information can be incorporated into the matching process allowing the framework to deal with vertex- and edge-attributed graphs in an uniform way. The experimental results showed that the approach is robust to structural and geometric noise, and capable of dealing with several relational abstractions of scene in a robust way. However, the scaling behavior of the association graph rendered the space requirements of our implementation very demanding. This severely limited the approach in dealing only with sparse graphs. Note, however, that holding the full edge-association graph explicitly in memory is not necessary. Future work include the reformulation as an implicit search on the edge-association space to make the approach usable with larger graphs.

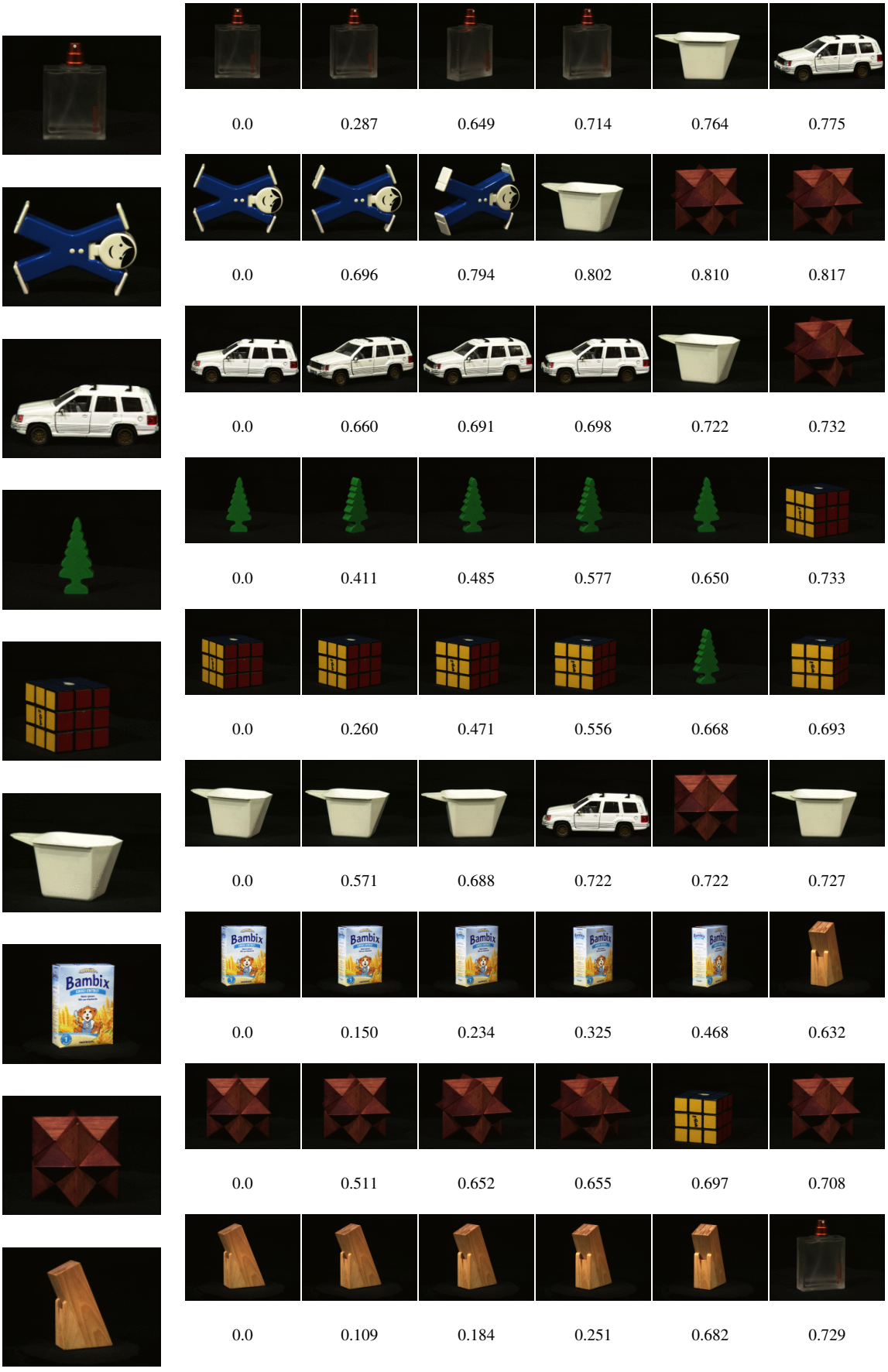


Figure 4.5: Retrieval with Delaunay graphs

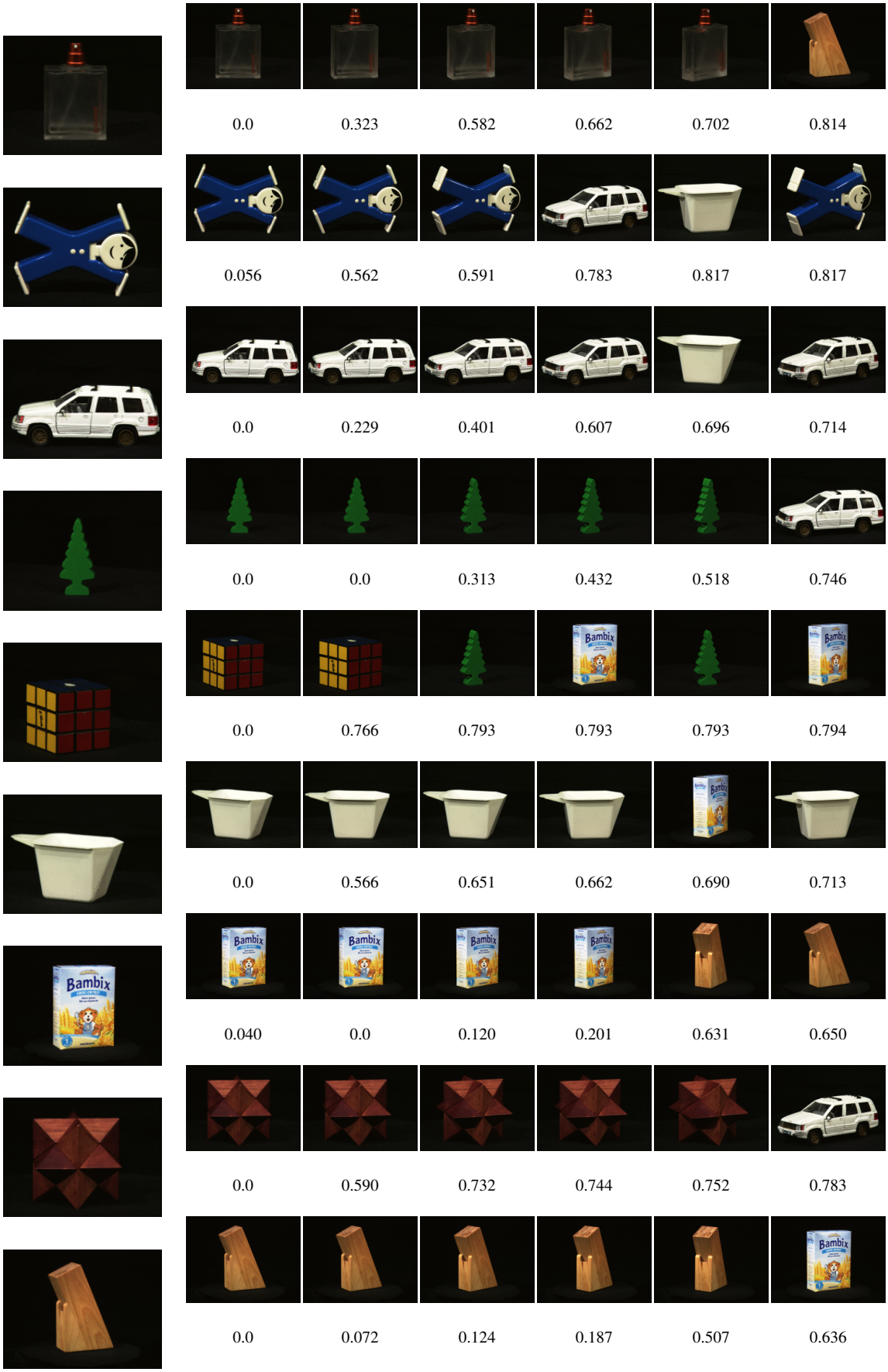


Figure 4.6: Retrieval with 5-neighbor graphs

4.2 Exploiting Structure to Match Articulated Objects

Object tracking with freely moving cameras is an open issue, since background information cannot be exploited for foreground segmentation, and plain feature tracking is not robust enough for target tracking, due to occlusions, distractors and object deformations. In order to deal with such challenging conditions a traditional approach, based on Camshift-like color-based features, is augmented by introducing a structural model of the object to be tracked incorporating previous knowledge about the spatial relations between the parts. Hence, an attributed graph is built on top of the features extracted from each frame and a graph matching technique based on Dominant Set clustering is used to find the optimal match with the model. Pixel-wise and object-wise comparison with other tracking techniques with respect to manually-obtained ground truth are presented.

4.2.1 Object Tracking and Matching

In recent years, object tracking has been recognized by the scientific community as a fundamental task in several applications of video analysis. Tracking rigid objects in simple, uncluttered scenes acquired from static cameras is an almost solved problem [181]. Conversely, in complex scenarios where objects camouflage with the background, have severe shape variations and are strongly occluded, tracking can be really challenging. In addition, when either the background is not fixed or the camera is moving, no statistical or geometrical model can be exploited to segment the foreground objects and predictive models (such as Kalman filters) are ineffective.

In point tracking, objects are usually represented by single or multiple points and the correspondences between two consecutive frames is established by either deterministic [171] or statistical methods [20] to provide tracking without object segmentation. An alternative is to represent the data using kernel primitives such as rectangles or ellipses. These kernel methods can be used to estimate a density-based appearance model of the object [49]. Other approaches encompass silhouette tracking, estimating the object contour evolution by means of state-space models [85] or variational methods [22].

These proposals are robust and efficient when the object can be represented by a single feature, such as the color histogram, but in the case of complex articulated objects represented by parts which are often partially or completely overlapped they are likely to fail. To deal with such challenging scenarios structural information expressing spatial constraint among features might be used. This is the case of the pictorial structures of [57] that have been proposed for object recognition and has been further developed for people tracking by [131]. Similarly [84, 150] are based on inference in a graphical model and can be applied again to people tracking [146]. All these approaches tend to be specifically focused on the articulated structure of the human body or human face (whereas our framework tackles generic-shape object tracking), and rely on Bayesian probabilistic frameworks; on the other hand tracking can be brought to a problem of graph matching through a graph based representation based on Region Adjacency Graph (RAG), where vertices represent image regions and edges encode adjacency. This is the case of [67, 50, 17, 69].

A notable exception is [86] where RAGs are tracked by fitting independent Kalman filters to both regions and adjacency relations. [140] uses graphs and Kalman filter for insects tracking.

Structural methods based on point features are less used than region-based ones. This is due primarily to the fact that it is more difficult to define relations between point features. In [155] SIFT features are extracted from the tracked object and a nearest-neighbor graph is built on top of them. Relaxation labelling is used for matching and the object graph itself is updated by removing disappearing features and adding new ones. In [29] the features tracked are the linear borders of geometric objects and edges connect parallel or perpendicular borders.

The definition of the structural model can be inferred from the image data. This approach is very general but might suffer from the instability of the model inference, both in terms of detection of regions/features and with respect to the invariance of the relational structure to be tracked. In addition an inferred model is inherently unable to capture detailed information about the intrinsic articulation and deformability of non-rigid objects. By contrast, our approach requires an a priori structural model of the target object (not necessarily bound to the human body figure), that is then enriched with attributes extracted from real data. This way we are able to search for a match that not only maximizes the coherence between attributes, but that also accounts for the coherence of the structural relations in a way invariant to variations in scale, rotations and translations, and even blurring due to camera motions; the search for the best coherent match with the provided model is made through Dominant Set extraction.

4.2.2 Overview of the Framework

Fig. 4.7 shows the conceptual scheme of our framework. An initial *Graph-Based Model Definition* provides the framework with both a model of the features to be tracked and a structural representation of their spatial arrangement. In this work color features are used, but different or more descriptive features (e.g. textures, edges) can be exploited. Moreover, an initial image can be used as reference for the extraction of the feature model (as in our current implementation), the structural model or both (Fig. 4.8a). Each new frame I_t is provided to the *Feature Cluster Extraction* component (Sect. 4.2.3) that applies the feature model and produces the mask of the probability of each feature class onto the current image (called *back-projection* [30]). Each back-projection is then clustered using meanshift and, for each cluster, attributes are extracted. Most of the extracted feature clusters represent erroneous detections of the tracked object feature (see Figs. 4.8b-f) and the correct candidates must be extracted using global consistency information.

A labelling function maps each feature cluster on the originating model feature. Each pair of clusters whose features are rigidly joined together in the structural model, are connected by edges to form the *labelled graph* G_t (Sect. 4.2.4). Then an edge weighted *association graph* GA_t is created between the structural model G_0 and the labelled graph G_t (Sect. 4.2.4) and each edge is weighted according to a global *coherence measure* (Sect. 4.2.4) in such a way that each maximal edge weight clique in GA_t corresponds to

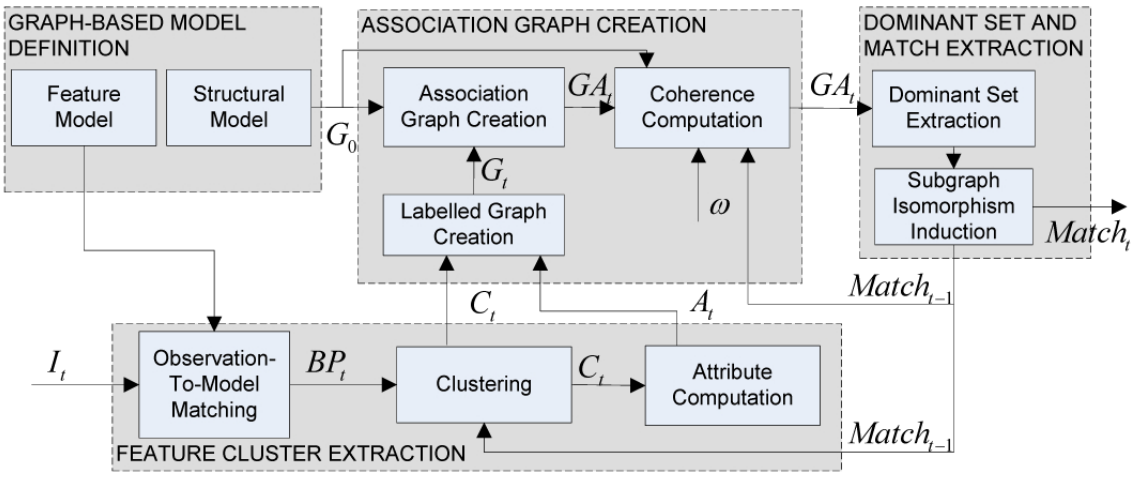


Figure 4.7: Scheme of the proposed framework.

a maximal coherence subgraph isomorphism and vice versa. Finally the Dominant Set framework is used to search for the maximum coherence $Match_t$ (Sect. 4.2.4, Fig. 4.8g).

4.2.3 Extraction of Feature Clusters

For each feature of the model, the *Feature Cluster Extraction* component extracts all the possible clusters of features which might represent a part of the tracked object according to the feature model. In our work, the *Feature Cluster Extraction* operates on simple color features using a modification of the Camshift algorithm [30], but different cluster extraction algorithms can be used.

The standard Camshift tracking algorithm uses a model of the object, consisting of a color histogram, and requires a region of interest to initialize the search. For each input image a probability mask of the model is produced, evaluating each pixel according to the color histogram as if it were a pdf. The resulting value is then scaled on 256 gray levels, producing the so called back-projection. Then, iteratively alternating the meanshift gradient ascend algorithm and a size-adaptation of the region of interest, the region estimate converges to encompass the extracted features and then provides the initial location for the next frame.

For the extraction of the feature clusters, the Camshift is modified as follows. *First*, the object to be tracked is modeled with multiple color histograms, corresponding to different areas of the objects (e.g. Fig. 4.8a); therefore, for each input image, multiple back-projections (BP_t) are obtained (Figs. 4.8b-f) and the cluster extraction proceeds independently for each BP_t. *Second*, the back-projections are obtained on the following color space:

$$(h,s,v) = \begin{cases} \left(\left\lfloor \frac{H}{16} \right\rfloor, \left\lfloor \frac{S}{16} \right\rfloor, \max_V\right) & \text{if } S > \tau_S \wedge V > \tau_V \\ \left(0, 0, \left\lfloor \frac{V}{16} \right\rfloor\right) & \text{otherwise} \end{cases} \tag{4.7}$$

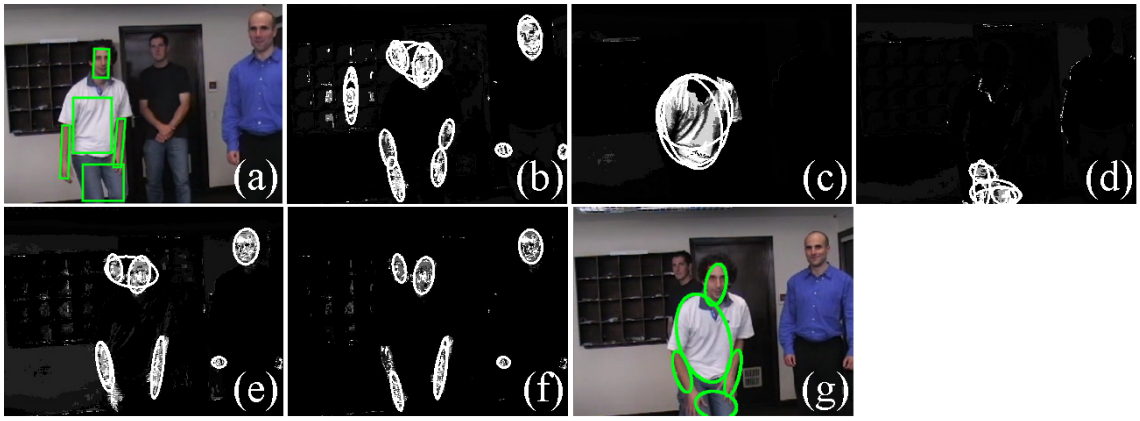


Figure 4.8: Framework steps example: (a) model to track, (b-f) feature back-projections and clusters, (g) best coherence match

The addition of value and saturation components to the standard Camshift color space allows us to deal better with low-saturation (considering only the V component) and provides an enriched color description. *Third*, our approach scatters particles over the BP_t from which to start the cluster extraction, producing therefore several clusters. The particles are spatially scattered over the BP_t with Gaussian or uniform distribution, depending on the object tracking status at the previous frame.

For each cluster C_t^i of the set C_t, the set of attributes $A_t^i = \left(D\left(C_t^i\right), M\left(C_t^i\right), P\left(C_t^i\right), R\left(C_t^i\right)\right)$ are computed, where D is the density, M the mass, $P = (x, y)$ the coordinates of the cluster's centroid and R the area:

$$M\left(C_t^i\right) = \sum_{p \in C_t^i} BP_t\left(p\right) \; ; R\left(C_t^i\right) = \left\| \left\{ p \in C_t^i \right\} \right\|$$
$$D\left(C_t^i\right) = M\left(C_t^i\right) / R\left(C_t^i\right) \tag{4.8}$$

4.2.4 Tracking using relational information

Regardless of the robustness of the extraction step several factors could lead to a wrong assignment between clusters. In fact, distractors, noise, deformation or pose and illumination changes can easily lower the coherence between correct correspondences or make unrelated features more similar. For this reason any approach that is based only on the similarity between features is inherently sensitive to noise. To overcome this limitation we add contextual information, thus casting the feature matching into a more robust subgraph matching problem.

From feature clusters to labelled graphs

In order to obtain a graph from a set of feature clusters we exploit the previous knowledge about the physical structure of the object. To this end, we define a structural model where

each part of the object is associated to a feature class which is known to be rigidly joined to some other parts, but can move freely from the rest. This is the case with any articulated object, while totally-rigid objects can be modeled by joining all the parts.

A *structural model* of an object is a connected graph $G_m = (P, S)$ where P is the set of distinct parts we want to use to represent the object and $S \subseteq P \times P$ are their structural relations, where $(p_a, p_b) \in S$ iff p_a and p_b are joined in the object. This model embeds our previous knowledge about the structure of the object to be tracked in terms of its parts. In Fig. 4.9 some examples of structural models are presented.

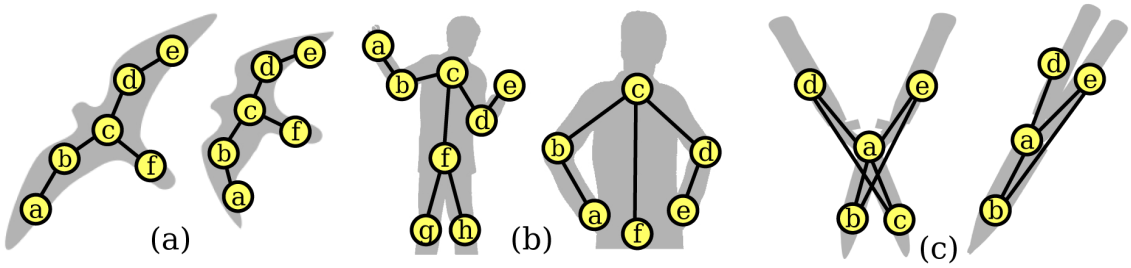


Figure 4.9: Example of structural models and labelled graphs. The model is subject to deformation (a) and also to scaling and occlusion (b); in (c) it comprises two totally-rigid submodels partially occluding each other

Given a structural model $G_m = (P, S)$, a set of features clusters C assigned to $|P|$ classes by a surjective labelling function $l : C \rightarrow P$ and their attributes A, we define the *labelled graph* as the $|P|$-partite graph $G = (C, E, A, l)$ where C is the vertex set, $E = \{(u, v) \in C \times C | (l(u), l(v)) \in S\}$ the edge set, A the vertex attributes and l the vertex labelling function. In this graph each edge represents a structural relation between a pair of feature clusters. The automatic extraction of feature cluster candidates from a frame I_t yields a graph with many nodes and edges. The supervised selection of the ground truth from a reference frame will result in a simpler graph with just one cluster for each part of the object to be tracked: we call this graph the model graph. Our goal is to find within each labelled graph extracted from a frame I_t the subgraph which is the most coherent with the model graph we are tracking. In other words we are looking for a maximum coherence subgraph isomorphism.

Given labelled graphs $G_1 = (C_1, E_1, A_1, l_1)$ and $G_2 = (C_2, E_2, A_2, l_2)$ a *labelled isomorphism* between them is a relation $M \subseteq C_1 \times C_2$ such that for each $(u_1, u_2), (v_1, v_2) \in M$ the following properties hold:

$$l_1(u_1) = l_2(u_2) \text{ and } l_1(v_1) = l_2(v_2) \tag{4.9}$$

$$u_1 = v_1 \Leftrightarrow u_2 = v_2 \tag{4.10}$$

The first condition ensures that M does not map feature cluster of incompatible classes. The second condition forces M to be a partial injective function. It is easy to see that any labelled isomorphism is a special case of subgraph isomorphism which enforces label consistency.

We still need to define a measure of the global coherence of a labelled isomorphism M. In our context limiting the measure to a similarity between vertex attributes would be not enough, as this way we would be unable to take into account structural relations among vertices. Unfortunately, even measuring coherence between edges would not be general enough, as it would not be possible to account for invariants that depends on more than one edge, such as length ratios or angle differences. For this reason we defined a coherence measure between pairs of edge matches as this allows us to deal with variations in scale and articulation throughout the video sequence. To this end we define the set of edge matches as:

$$\begin{aligned} e(M) = \{((u_1, v_1), (u_2, v_2)) \in E_1 \times E_2 | \\ (u_1, u_2) \in M \wedge (v_1, v_2) \in M\} \end{aligned} \tag{4.11}$$

and let $\omega : (E_1 \times E_2) \times (E_1 \times E_2) \rightarrow \mathbb{R}^+$ be a measure of coherence between pairs of edges matches, then the total weight of M is defined as:

$$\Omega(M) = \sum_{a \in e(M)} \sum_{b \in e(M) \setminus \{a\}} \omega(a, b)\,. \tag{4.12}$$

From graph matching to clique search

In order to search for a match of maximum compatibility between two labelled graphs we choose a two-step approach which first casts the matching problem into a clique search problem and then solves it using continuous optimization.

Given labelled graphs $G_1 = (C_1, E_1, A_1, l_1)$ and $G_2 = (C_2, E_2, A_2, l_2)$ and a function $\omega : (E_1 \times E_2) \times (E_1 \times E_2) \rightarrow \mathbb{R}^+$ that measures the coherence between pairs of edge associations, we define an association graph between them as an edge weighted graph $Ga = (Va, Ea, \omega)$ where $Va = E_1 \times E_2$, $Ea \subset Va \times Va$ with $(((u_1, v_1), (u_2, v_2)), ((w_1, z_1), (w_2, z_2))) \in Ea$ iff:

$$l_1(u_1) = l_2(u_2), l_1(v_1) = l_2(v_2),$$

$$l_1(w_1) = l_2(w_2) \text{ and } l_1(z_1) = l_2(z_2) \tag{4.13}$$

$$u_1 = w_1 \Leftrightarrow u_2 = w_2, v_1 = z_1 \Leftrightarrow v_2 = z_2 \tag{4.14}$$

With this definition we are able to show some useful connections between labelled isomorphisms and complete subgraphs (cliques) in this association graph.

To this end, note that each $X \subseteq V_a$ represents a relation between edges in E_1 and E_2. In order to obtain a relation between vertices in V_1 and V_2 we define a natural map $v : \mathcal{P}(V_a) \rightarrow \mathcal{P}(V_1 \times V_2)$ as:

$$\begin{aligned} v(X) = \{(u_1, u_2) \in V_1 \times V_2 | \\ ((u_1, v_1), (u_2, v_2)) \in X \vee ((v_1, u_1), (v_2, u_2)) \in X\} \end{aligned} \tag{4.15}$$

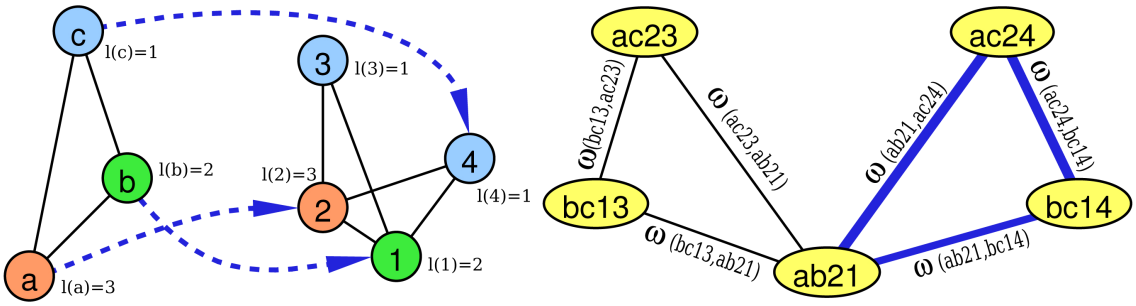


Figure 4.10: Labelled isomorphism between two labelled graphs and the clique associated to it in the edge weighted association graph.

That is, a match between vertices is induced by X if they are mapped by any edge match in X. It is easy to see that v is not injective, nevertheless it has a proper right partial inverse, namely the function $e(M)$ defined by (4.11).

We now formulate the following lemmas:

Lemma 1. *Given labelled graphs G_1, G_2 and their association graph G_a, $X \subseteq V_a$ is a clique iff $v(X)$ is a labelled isomorphism between G_1 and G_2.*

Proof. We will prove both the implications by contrapposition. First we will prove that if X is not a clique then $v(X)$ is not a labelled isomorphism:

Suppose that X is not a clique. This means that it contains at least a pair of vertices $((u_1, v_1), (u_2, v_2)) \in X$ and $((w_1, z_1), (w_2, z_2)) \in X$ that do not satisfy either (4.13) or (4.14). Those vertices induce associations (u_1, u_2), (v_1, v_2), (w_1, w_2) and (z_1, z_2) in $v(X)$. Suppose that (4.13) is not satisfied, this means that some label correspondence is not satisfied. For instance, this happens if $l(u_1) \neq l(u_2)$, thus in this case (4.9) is not satisfied by the induced association (u_1, u_2). The same happens if $l(v_1) \neq l(v_2)$, $l(w_1) \neq l(w_2)$ or $l(z_1) \neq l(z_2)$. Suppose that (4.14) is not satisfied, this means that some injection condition is not satisfied. For instance, this happens if $u_1 = w_1$ and $u_2 \neq w_2$, this in this case (4.10) is not satisfied by the induced associations (u_1, u_2) and (w_1, w_2). The same happens for the others conditions. In both cases $v(X)$ is not a label isomorphism, which proves our claim.

Now we will prove that if $v(X)$ is not a labelled isomorphism between G_1 and G_2 then X is not a clique:

Suppose that $v(X)$ is not a label isomorphism between G_1 and G_2. This means that it contains a pair of associations $(u_1, u_2) \in v(X)$ and $(w_1, w_2) \in v(X)$ that do not satisfy either (4.9) or (4.10). For such matches to be in $v(X)$ it means (by Def. 4.15) that either $((u_1, v_1), (u_2, v_2)) \in X$ or $((v_1, u_1), (v_2, u_2)) \in X$ for some $v_1 \in V_1$ and $v_2 \in V_2$ and that either $((w_1, z_1), (w_2, z_2)) \in X$ or $((w_1, z_1), (w_2, z_2)) \in X$ for some $z_1 \in V_1$ and $z_2 \in V_2$. Suppose that (4.9) is not satisfied because $l(u_1) \neq l(u_2)$ and that $((u_1, v_1), (u_2, v_2)) \in X$,

in this case neither (4.13) is satisfied and $((u_1, v_1), (u_2, v_2))$ is not adjacent to any other vertex in G_a, thus X is not a clique in G_a. Other cases that do not satisfy (4.9) are similar. Similar observations show that if (4.10) is not satisfied a pair of vertices not connected by an edge exist in X. In both cases X is not a clique in G_a, which proves our claim.

□

Lemma 2. *If $X \subseteq V_a$ is a maximal clique in G_a, then $v(X)$ is a maximal labelled isomorphism between G_1 and G_2. Conversely, if M is a maximal labelled isomorphism between G_1 and G_2 then $e(M)$ is a maximal clique in G_a.*

Proof. We will prove both the implications by contrapposition. First we will prove that if $v(X)$ is not a labelled isomorphism then X is not a maximal clique:

Suppose that $v(X)$ is a labelled isomorphism between G_1 and G_2 but it is not maximal. This means that the association (u_1, u_2) can be added to $v(X)$ without invalidating the conditions (4.9) or (4.10). We know from the definition of structural model that the model is connected and from the definition of labelled graph that for each part in the model at least one vertex in G_1 and G_2 is associated to the corresponding label by l. This guarantees that at least $(u_1, v_1) \in E_1$, $(u_2, v_2) \in E_2$ or $(v_1, u_1) \in E_1$, $(v_2, u_2) \in E_2$ for some v_1 and v_2 with $l(v_1) = l(v_2)$. We know form Lemma 1 that X is a clique in G_a. In addition, in either cases the associations between those edges can be added to X without invalidating (4.13), as the labelling correspondence for u_1,u_2 is guaranteed and v_1, v_2 can be chosen of the same class. Also (4.14) will not be invalidated by the new associations, as if v_1 is not yet in X any compatible v_2 will not invalidate it, otherwise the already present v_2 can be chosen. This shows that X is not maximal, which prove our claim.

Now we will prove that if $e(M)$ is not a maximal clique then M is not a maximal labelled isomorphism.

Suppose that $e(M)$ is a clique in G_a but it is not maximal. Since $e(M)$ contains all the edge associations between correspondent vertices in M if a new edge association can be added to it without invalidating (4.13) and (4.14), then it induces new vertices associations in M that do not invalidate (4.9) and (4.10). Thus M is not maximal, which prove our claim. □

From the previous lemmas and the definition of the weight of a labelled isomorphism M, derives the following:

Theorem 4. *Given two feature graphs G_1 and G_2, each maximal(maximum) weight labelled isomorphism M between them induces a maximal(maximum) edge weight clique in $Ga(G_1, G_2)$ and vice versa.*

Proof. Lemma 2 proves the maximality correspondence. In addition, the definition of $e(M)$ states that the weight of the label isomorphism M is exactly the sum of the weights assigned to the edges of $e(M)$ by the similarity function ω. Thus, for each maximal(maximum)

labelled isomorphism M a maximal(maximum) edge weight clique $e(M)$ of exactly the same weight exists. In addition, it is easy to see that if X is maximal then $X = e(v(X))$, thus for each maximal(maximum) edge weight clique X in Ga a maximal(maximum) labelled isomorphism $v(X)$ of exactly the same weight exists. □

Fig. 4.10 shows an example of a labelled isomorphism and the correspondent clique in a labelled association graph.

An effective heuristic for the weighted clique problem

Theorem 4 casts our tracking problem into a search for a maximal edge weighted clique in a novel type of association graph. In order to perform this search we use the Dominant Set framework [123]. Given an edge weighted graph $G = (V, E, \omega)$, a subset of vertices $S \subseteq V$ and two vertices $i \in S$ and $j \notin S$, the following function measures the coherence between nodes j and i, with respect to the average coherence between node i and its neighbors in S:

$$\phi_S(i,j) = \omega(ij) - \frac{1}{|S|}\sum_{k \in S} \omega(ik) \tag{4.16}$$

While overall weighted coherence between i and all the nodes in S is defined as:

$$w_S(i) = \begin{cases} 1 & \text{if } |S| = 1 \\ \sum_{j \in S\setminus\{i\}} \phi_{S\setminus\{i\}}(i,j) w_{S\setminus\{i\}}(j) & \text{otherwise} \end{cases} \tag{4.17}$$

Intuitively, $w_S(i)$ will be high if i is highly coherent with vertices in S. Given this measure $S \subseteq V$ is said to be *dominant* if the following conditions hold:

$$w_S(i) > 0, \forall i \in S \text{ and } w_{S\cup\{i\}}(i) < 0, \forall i \notin S \tag{4.18}$$

The conditions above correspond to the two main properties of a cluster: namely internal homogeneity and external inhomogeneity. In the literature this framework has been associated to clustering, nevertheless its use as an heuristic for the edge weighted clique problem is justified by the fact that, when applied to unweighted graphs, the notion of a dominant set is equivalent to the notion of a clique. Hence, a dominant set can be seen as a generalization of cliques to graphs with weighted edges. Moreover there is another compelling reason to prefer dominant sets over traditional techniques of clique search: in fact their clustering property allows us to discard automatically nodes that are less coherent with respect to the others. This is the case when a part of the model is missing or occluded. For instance in Fig. 4.11 the face is out of the frame border, but candidates for it are generated anyway by the back projection: in this situation an exact graph matching would wrongly include in the result also the best of those candidates (green ellipse), whereas dominant sets leave it out as its coherence is very low with respect to the other parts in the result (red ellipses). It is worth noting that this selection does not require the user to choose a threshold as it is implicit in the cluster properties.

Pavan and Pelillo [123] have shown that dominant sets correspond to local maximizer over the standard simplex of the quadratic function $f(\mathbf{x}) = \mathbf{x}^t A\mathbf{x}$ where A is the weighted adjacency matrix of the graph (thus $A_{ij} = \omega(i,j)$). These maximizers can be found by exploiting the convergence properties of the payoff monotonic replicator dynamic $x_i(t+1) = x_i(t)(Ax(t))_i/(x(t)^t Ax(t))$ which is guaranteed to converge to a local maximum when the association graph is undirected and, thus, the matrix A is symmetric [126]. At convergence the value of the function f is a measure of the coherence of the extracted set. This property is used to detect the absence of the object from the scene and suspend the tracking. Finally, as the local maximizer found by the replicator dynamic is not guaranteed to be the global maximum, we used an enumeration strategy similar to the one presented in [135].

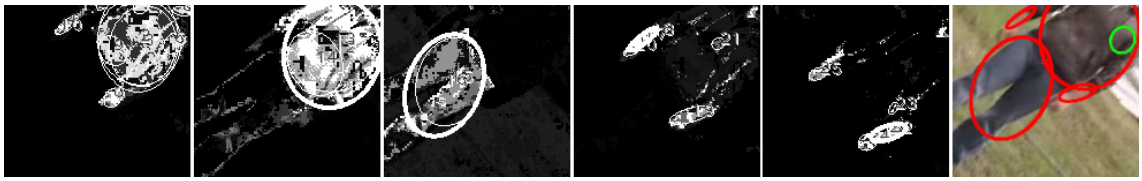


Figure 4.11: Example of the failing of an exact graph matching

Coherence Computation

Given the association graph $Ga_{t,0}$ between G_t and G_0, our goal is to assign to each of its edges $(((u_t, v_t), (u_0, v_0)), ((w_t, z_t), (w_0, z_0))) \in Ea_{t,0}$ a weight in the interval $[0, 1]$ which reflects the coherence between the two connected edge associations (see Fig. 4.10). This measure $\omega : Ea_{t,0} \rightarrow [0, 1]$ is the sum of several components, each referring to a specific property of the tracked object that should be consistent along the video sequence. Since different and independent properties are considered, the mis-detection of any of them (for example, due to occlusion or deformation) does not compromise the overall coherence evaluation. In the present work we define three properties that are expected to be consistent along the video sequence: *color* and *structure* w.r.t. the initial model, and *spatial similarity* w.r.t. the previous frame.

Color-based consistency measured through cluster density ω_d and mass ω_m: let us define the normalized density and the normalized mass respectively as:

$$ND(u_t) = \frac{D(u_t)}{\max\limits_{\forall v_t \in C_t \mid l(u_t)=l(v_t)} D(v_t)} \tag{4.19}$$

$$NM(u_t) = \frac{M(u_t)}{\max\limits_{\forall v_t \in C_t \mid l(u_t)=l(v_t)} M(v_t)} \tag{4.20}$$

then:

$$\omega_d = \sqrt[4]{ND(u_t) \cdot ND(v_t) \cdot ND(w_t) \cdot ND(z_t)} \tag{4.21}$$

$$\omega_m = \sqrt[4]{NM(u_t) \cdot NM(v_t) \cdot NM(w_t) \cdot NM(z_t)} \tag{4.22}$$

The clusters are defined over the back-projection that measures the color similarity of the image I_t compared to a color feature of the model: therefore the higher the density of a cluster, the higher its color similarity to the model. The densities of the four clusters are multiplied and not summed up in order to reinforce the overall $Ea_{t,0}$ color similarity. Since small clusters might show very high ω_d, the ω_m component reinforces only the $Ea_{t,0}$ that have strong masses.

Structure consistency measured through cluster sizes and inter-cluster distances ω_{sd}: this component reinforces the $Ea_{t,0}$ that shows structural similarity with the model, i.e. cluster size variations which are supported by consistent inter-cluster distance variations. Fig. 4.12 depicts three different cases. (a) is a typical structure size reduction (for example, due to camera zoom out) that maintains consistency between area and distance variations. On the other hand, (b) and (c) depict a structure deformation that is penalized by ω_{sd}: in both cases the distance variation between top and middle clusters is not supported by a similar variation in the size of the cluster; ω_{sd} is formalized introducing the *linear area ratio* and the *distance ratio* respectively as:

$$lar : C_t \times C_0 \rightarrow [0, \infty), lar\left(u_t, u_0\right) = \sqrt{\frac{R\left(u_t\right)}{R\left(u_0\right)}}$$

$$dr : E_t \times E_0 \rightarrow [0, \infty), dr\left(\left(u_t, v_t\right), \left(u_0, v_0\right)\right) = \frac{\left|\overline{P\left(u_t\right) P\left(v_t\right)}\right|}{\left|\overline{P\left(u_0\right) P\left(v_0\right)}\right|}$$

Structure consistency of $Ea_{t,0}$ is obtained when lar measures are similar to the respective dr measures, i.e. their ratio is close to 1; the consistency measure can then be obtained modeling the deviation with a Gaussian. To evenly stretch the ratio codomain from $[0, \infty)$ to $(-\infty, \infty)$, it is appropriate to compute the logarithm. Therefore, ω_{sd} is defined as follows:

$$\begin{aligned}\omega_{sd} =& e^{\frac{-(Q(u)-\Delta(u,v))^2}{2\sigma^2}} \cdot e^{\frac{-(Q(v)-\Delta(u,v))^2}{2\sigma^2}} \cdot \\ & \cdot e^{\frac{-(Q(w)-\Delta(w,z))^2}{2\sigma^2}} \cdot e^{\frac{-(Q(z)-\Delta(w,z))^2}{2\sigma^2}}\end{aligned} \tag{4.23}$$

where $Q\left(a\right) = \log\left(lar\left(a_t, a_0\right)\right)$ and $\Delta\left(b, c\right) = \log\left(dr\left(\left(b_t, c_t\right), \left(b_0, c_0\right)\right)\right)$. In analogy to what is done with ω_d and ω_m, the four contributes of ω_{sd} are multiplied together and not summed up.

Structure consistency measured through cluster relative orientations ω_a : this component favors the maintenance of angular consistency of the $Ea_{t,0}$. Fig. 4.12 depicts two cases: regardless of the *overall* rotation of one graph compared to the other, (d) maintains the consistency of reciprocal angles of the segments, while (e) does not and is therefore penalized by ω_a. Let $\vartheta\left(\left(u_t, v_t\right), \left(u_0, v_0\right)\right) = \arccos \frac{\overline{P(u_t)P(v_t)} \times \overline{P(u_0)P(v_0)}}{\left\|\overline{P(u_t)P(v_t)}\right\| \cdot \left\|\overline{P(u_0)P(v_0)}\right\|}$ as the angle between the two segments connecting the centroids of the clusters, the value ω_a is defined as:

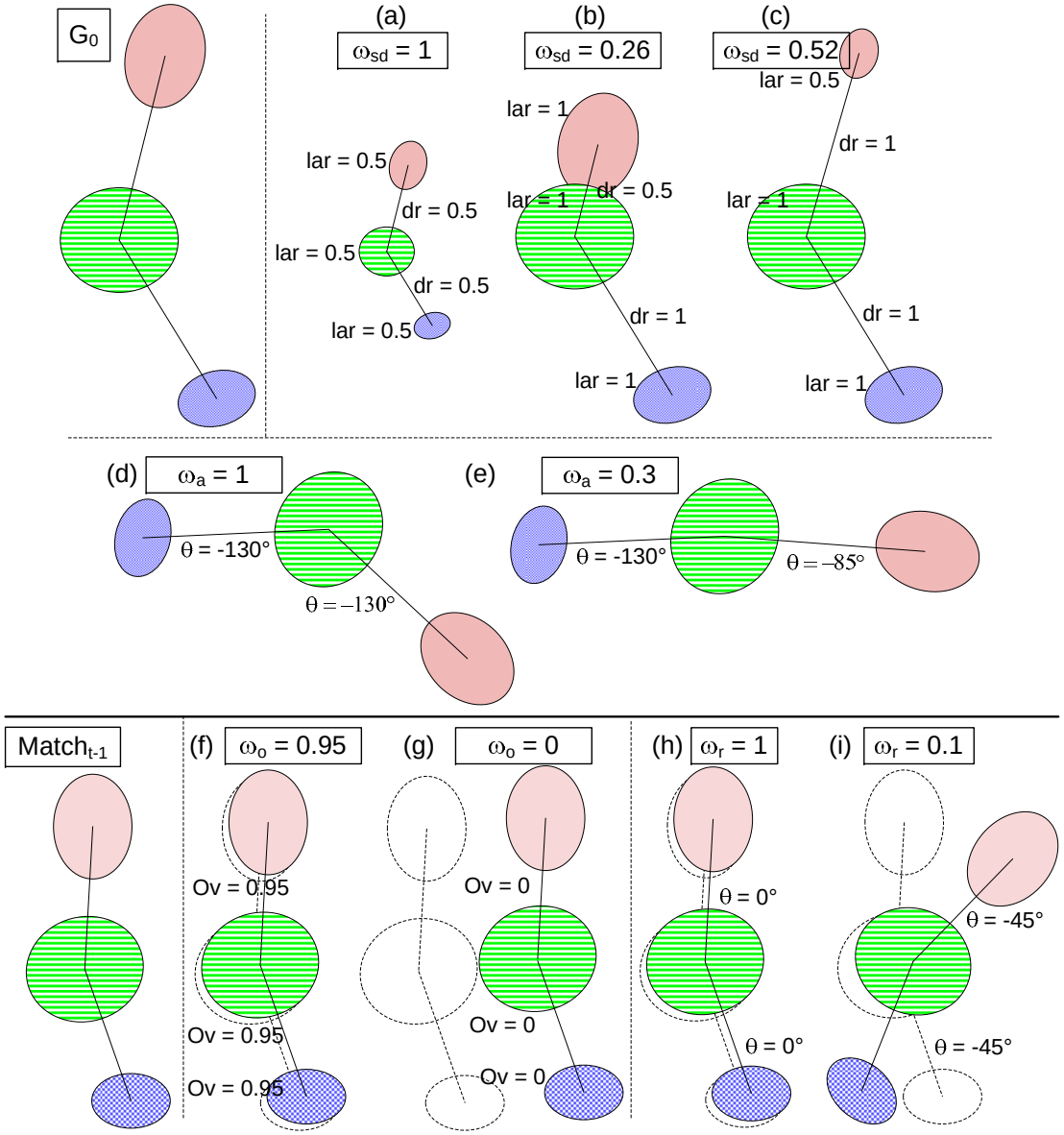


Figure 4.12: Structure consistency measure with ω_{sd} (a, b, c) and ω_a (d, e), and spatial similarity measure with ω_o (f, g) and ω_r (h, i). For the only sake of clarity and without loss of generality, the G_0 and $Match_{t-1}$ are made of only three nodes.

$$\omega_a = \frac{\exp\left\{m \cdot \cos\left[\vartheta\left(\left(u_t, v_t\right), \left(u_0, v_0\right)\right) - \vartheta\left(\left(w_t, z_t\right), \left(w_0, z_0\right)\right)\right]\right\}}{\exp\left\{m\right\}} \tag{4.24}$$

This resembles the Von Mises distribution [109], that is often used to model angular distributions.

Spatial similarity with the object at previous frame measured through overlap ω_o and rotation ω_r: let us consider the graph $Match_{t-1}$, which represents the tracked object at the previous frame, and the projection of its attributes over the graph G_0; the similarity components ω_o and ω_r favor the edges in $Ea_{t,0}$ that respectively maximize the area of overlap and minimize the overall graph rotations with respect to $Match_{t-1}$. In case $Match_{t-1}$ is partial/missing, these components will provide the contribution for the detected portion of the object only. Fig. 4.12 (f,g,h,i) depicts some explanatory examples. By defining $Ov\left(u_t, u_{t-1}\right) = \frac{2 \cdot R(u_t \cap u_{t-1})}{R(u_t)+R(u_{t-1})}$, we have:

$$\omega_o = \sqrt[4]{Ov\left(u_t, u_{t-1}\right) Ov\left(v_t, v_{t-1}\right) Ov\left(w_t, w_{t-1}\right) Ov\left(z_t, z_{t-1}\right)} \tag{4.25}$$

ω_r is defined to favor the minimization of the rotation of each single segment:

$$\begin{aligned} \omega_r = & \frac{\exp\left\{m \cdot \cos\left[\vartheta\left(\left(u_t, v_t\right), \left(u_{t-1}, v_{t-1}\right)\right)\right]\right\}}{\exp\left\{m\right\}} \cdot \\ & \cdot \frac{\exp\left\{m \cdot \cos\left[\vartheta\left(\left(w_t, z_t\right), \left(w_{t-1}, z_{t-1}\right)\right)\right]\right\}}{\exp\left\{m\right\}} \end{aligned} \tag{4.26}$$

4.2.5 Experimental Results

To demonstrate the advantage that the proposed graph based tracking (GB from now) offers with respect to existing techniques, we selected two tracking algorithms to compare with: Camshift (CS from now) and a *particle filtering* tracking based on color features (PF from now) similar to that proposed in [119]. For the sake of a fair comparison all three approaches are applied to the same color space (4.7).

In contrast to our approach, CS and PF do not correlate the results on the different feature models, that is, they do not exploit structure model. Therefore, we issue several independent instances of the algorithm on each single feature of the same object model used for the GB. They work well in standard conditions, but for the intrinsic limitation due to the lack of a structure model, they are likely to fail in challenging conditions, especially in the case of occlusions.

Our test bed consists of 3 videos and in one of them (Video 3) the tracking is applied twice, on two different target objects[1]. Table 4.1 summarizes the main characteristics of the benchmark videos.

In order to evaluate the performance of the approaches, we manually extracted the ground truth with the help of the VIPER-GT tool [2], consisting of several oriented bounding boxes, each containing a single part of the object to be tracked. Given the ground truth and the output of the tracking algorithms, it is possible to automatically compute the performance based on a set of metrics. Specifically, using the VIPER-PE tool [2], we obtained true positives, false negatives, false positives and, from them, *recall* and *precision*; these measures were extracted both at object and pixel level.

[1] Downloaded by AVSS 2007 dataset: ftp://motinas.elec.qmul.ac.uk/pub/multi_face

[2] http://viper-toolkit.sourceforge.net/

	Video 1	Video 2	Video 3 - a and b	
Generic info	Outdoor, moving cam, 1 person	Outdoor, moving cam, 2 persons	Indoor, static cam, 3 occluding people	
Model	F,T,P, LA,RA	F,T,P	F,T, H,P	F,T, LA,RA
Challenges	Severe scale vars and rotations, camouflaging background	Scene cuts, total obj. disapp, scale vars, rotations, camouflaging bkg	severe occlusions, several color distractors	

Table 4.1: Benchmark (F=face, T=torso, H=hands, P=pants, LA,RA= left/right arm).

The pixel-wise evaluation is shown in Fig. 4.2.5. In this case, we directly plot the frame-by-frame *F-measure* defined as an aggregation of recall R and precision P: $F = \frac{2 \cdot R \cdot P}{R+P}$. Pixel-wise recall and precision aggregate together the pixel measures performed separately on each single model class. $F = 1$ could reveal either a perfect matching (never happened in our tests), or the correct tracking suspension when the whole object is absent from the scene. Conversely, $F = 0$ reveals either a total failure or the detection of an object when this is not present.

Table 4.2 reports the summary of the pixel-wise and object-wise performance on the benchmark videos. Differently from the pixel-wise evaluation that merges together the pixel evaluations of all the tracked model classes, the object-wise evaluation gives a fairer evaluation on the tracking of the single classes, regardless of their pixel areas (e.g. it equally weights the tracking of a small part like a hand as the tracking of a bigger part like a torso). Since video 1 does not contain severe occlusions, scene cuts or object disappearances, the structural model in our approach does not significantly increase the performance compared to CS or PF, with exception of frames 199 and 231, when the face exits from the view: in fact, our approach correctly suspends the face tracking to resume it when the face reappears, whereas the other approaches fail. On the other hand, the sharp scene cuts (frames 156 and 231) and the full object disappearance (frame 156) of video 2 make the performance of CS and PF drop severely. Our approach instead is not affected at all, suspending the tracking when necessary and resuming it as soon as the structure of the model is found again.

Conversely, Video 3 is a static camera sequence but the persons occlude each other several times and the scene is full of color distractors (e.g. the several skin-colored regions of faces and arms, the two blue jeans, the dark t-shirt of the person on the right and the dark cupboard on the back wall). In such conditions the use of a structural model is determinant to have a successful tracking. As can be seen in sequence 3-a, our approach takes a few frames to resume the tracking since it needs to locate the structure first.

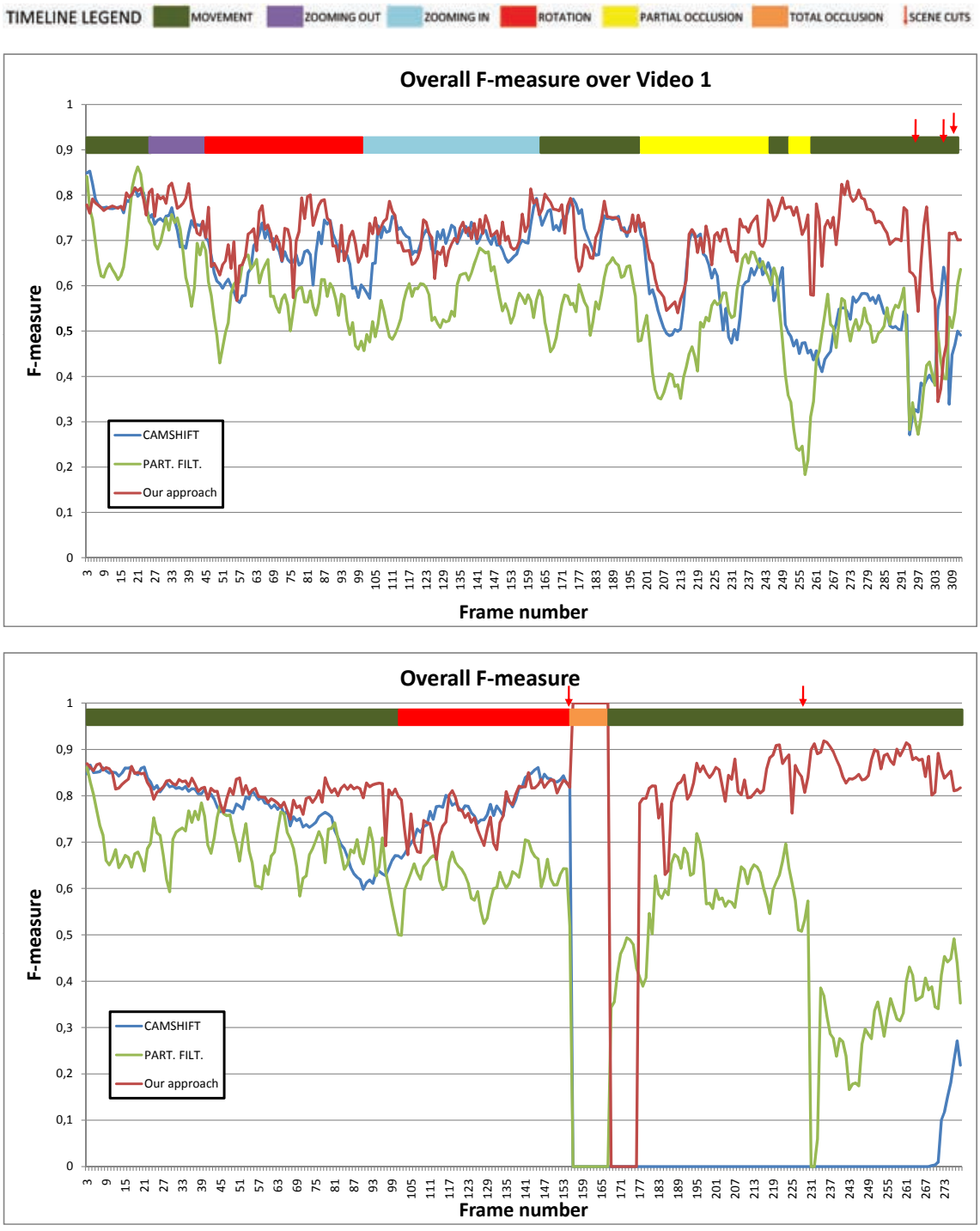


Figure 4.13: Pixel-level measure of performance. On top of each graph, a time line represents the different challenges on the tracking. The legend is on the top of the figure.

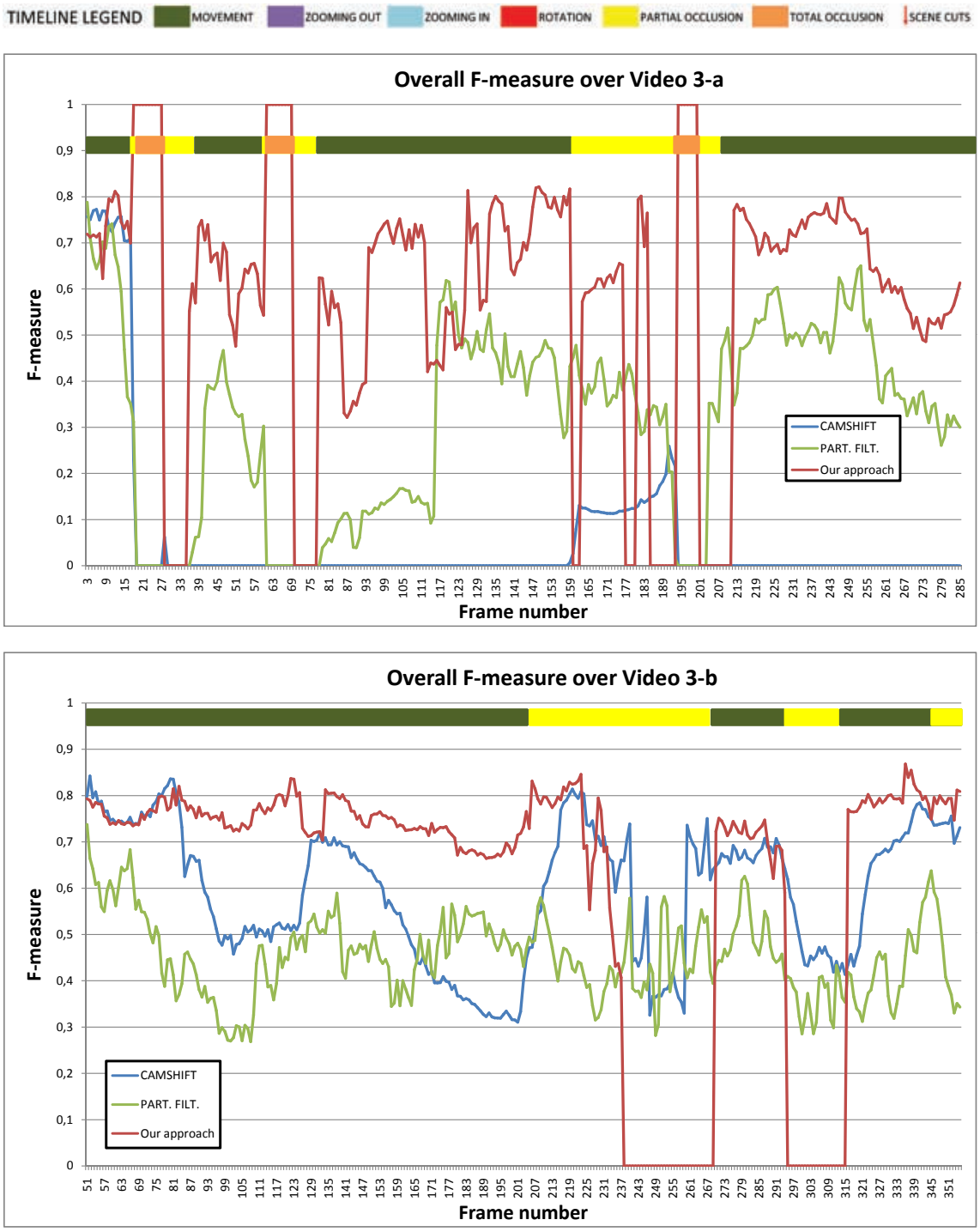


Figure 4.14: Pixel-level measure of performance. On top of each graph, a time line represents the different challenges on the tracking. The legend is on the top of the figure.

Object level	Recall			Precision			F-measure		
	CS	PF	GB	CS	PF	GB	CS	PF	GB
Video 1	96,72%	96,41%	99,24%	87,66%	74,52%	95,31%	91,97%	84,07%	97,24%
Video 2	92,83%	95,99%	100,00%	66,86%	89,00%	97,20%	77,73%	92,36%	98,58%
Video 3-a	30,25%	72,46%	97,87%	12,94%	69,89%	89,58%	18,13%	71,15%	93,54%
Video 3-b	88,59%	86,28%	98,13%	91,55%	79,39%	87,65%	90,05%	82,69%	92,59%
Pixel level	**Recall**			**Precision**			**F-measure**		
	CS	PF	GB	CS	PF	GB	CS	PF	GB
Video 1	84,54%	66,92%	85,71%	52,77%	49,80%	62,14%	64,09%	55,41%	71,16%
Video 2	53,20%	49,42%	84,87%	36,85%	64,64%	76,20%	43,34%	55,50%	79,78%
Video 3-a	7,93%	30,68%	65,26%	4,76%	36,93%	53,42%	5,67%	32,07%	57,67%
Video 3-b	65,06%	44,20%	71,95%	69,67%	74,30%	63,65%	64,66%	52,74%	66,90%

Table 4.2: Summary of the performance.

4.2.6 Concluding Remarks

The proposed joint feature-structure approach to object tracking in freely moving camera scenarios has shown interesting and promising results. In particular the experimental results show that the use of the structural approach give robustness to the tracking in the presence of severe occlusions and distractors. The coherence measure, used for weighting the association graph, is also used as a metric for the reliability of the tracking, allowing it to be suspended in case the object is not found in the scene. Moreover, the exclusion of low-coherence nodes from the extracted dominant set allows to reject false positive detections, often due to distractors. It is worth noting that the use of color features presented in this work is not a limitation, since the framework is flexible and open to be extended to different types of features. In fact, in future work we aim to include texture and edge features as well. Regarding the graph matching step, other search heuristics can be plugged into the framework in substitution of dominant sets. An extensive evaluation of the results obtained using different algorithms could be useful to choose the best performing technique in a general scenario. It should be noted that the current implementation of the dominant sets does not fit a real-time tracking as their extraction time for a single frame can span from one to several seconds, whereas discrete matching techniques could be much faster.

4.3 Using Hypergraphs to enforce Higher-Order Constraints

A problem commonly encountered in Computer Vision is the recovery of the transformation parameters between two affinely distorted images. In this section, we propose a novel feature-based approach that casts the matching problem to the search of a maximum clique over an auxiliary hypergraph. We also introduce a continuous-based characterization of cliques in hypergraphs that allows us to handle the hard combinatorial problem using tools from the continuous domain. Finally, we present experimental result and comparisons with a state-of-the-art algorithm.

4.3.1 Features and Parameter Estimation

The estimation of the transformation parameters between two affinely distorted images is a problem that is commonly encountered in many areas of Computer Vision. We can distinguish mainly two type of approaches: image-based and feature-based. The image-based approaches try to find a transformation that maximizes the overlap between the two images, usually by analyzing them in the frequency domain [25, 90, 105]. Conversely, feature-based approaches are characterized by two phases: Initially a set of features is extracted from each image and they are then matched to estimate the affine transformation [68, 102, 180, 179]. In the literature, we find several feature detectors [116] and descriptors [115] that can be employed for the first phase. Among them the most widely used is the Scale-Invariant Feature Transform (SIFT) [102]. After the two sets of features have been extracted, a correspondence between them is established, and assuming that their coordinates are related by a parametric transformation, the system of equations defined by them is solved in the least square error sense to derive the transformation parameters. Unfortunately, the set of features found by real world detectors are not perfectly conserved under affine transformation, and also descriptors can lead to false matching candidates. Thereby, there is the need of filtering out wrong transformations and generally this process is accomplished by means of a generalized Hough transform, where the model parameters are quantized into bins, and each extracted correspondence between features votes for the best transformation.

The method presented in this section falls in the class of feature-based approaches and focuses particularly on the solution of the second phase. A drawback of using a clustering mechanism to isolate the best affine transformation, is that the number of clusters is unknown a priori and a quantization of the parameter space can lead to imprecise solutions if the bins are too large, and to the dispersion of votes if they are too small. The solution that we propose overcomes these problems by casting the feature matching problem for the extraction of the affine transformation into a clique problem over an auxiliary hypergraph where the vertices correspond to feature associations and hyperedges correspond to groups of four associations that agree, within a desired tolerance, to the same affine transformation. In this way a maximum clique represents the largest group of features that agree on the same transformation. Since the maximum clique problem on hypergraphs is a relatively unexplored topic, a contribution is also to provide a continuous-based approach for it.

4.3.2 A Hypergraph Consistency Model

Before going into the details of the proposed approach, we need to introduce some notations and definitions regarding hypergraphs. Let A be a set and n a positive integer, with $\binom{A}{n}$ we denote the set of subsets of A of cardinality n. A *k-uniform hypergraph*, or simply a *k-graph*, is a pair $G = (V, E)$, where $V = \{1, \ldots, n\}$ is a finite set of *vertices* and $E \subseteq \binom{V}{k}$ is a set of *hyperedges*. Note that the concept of k-uniform hypergraph generalizes that of undirected graphs, in fact graphs are in a one-to-one relation with 2-graphs.

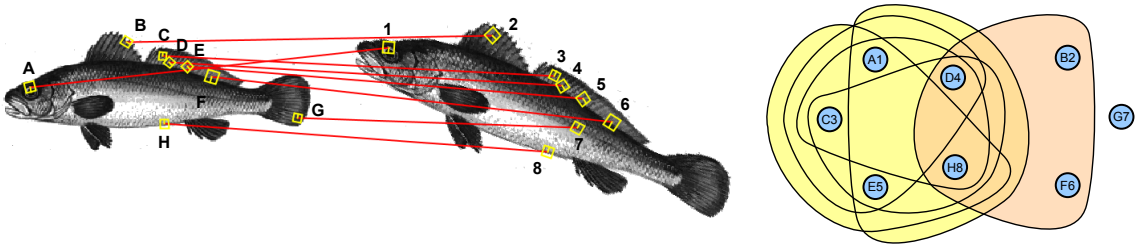


Figure 4.15: Example of auxiliary 4-graph for affine parameter estimation

The *complement* of a k-graph G is given by $\bar{G} = (V, \bar{E})$ where $\bar{E} = \binom{V}{k} \setminus E$. A subset of vertices $C \subseteq V$ is called a *clique* of G if $\binom{C}{k} \subseteq E$. A clique is said to be *maximal* if it is not contained in any other clique, while it is called *maximum* if it has maximum cardinality. We denote with $\mathcal{K}_n^k$ a complete k-graph with n vertices.

Given two sets of Euclidean features F_1 and F_2 extracted from the images, we build an auxiliary 4-graph $G = (V, E)$ where vertices are associations from $F_1 \times F_2$ and edges are groups of four correspondences that agree on the same affine transformation up to a given tolerance. In this way, a set of image features that agree, up to the specified tolerance, to the same affine transformation, form a clique of G. Thereby the problem of finding a robust set of correspondences that can be used to estimate the transformation is reduced to the problem of finding a clique on the auxiliary 4-graph with maximum cardinality. The method adopted to decide whether a set of four correspondences e agrees on a single affine transformation can be defined in many ways. Our solution consists in computing for every association $(\mathbf{x}, \mathbf{y}) \in e$ the affine transformation $(A, \mathbf{b})$ obtained from the remaining 3 associations and calculating the transformation error on $\mathbf{x}$ given by $\|A\mathbf{x} + \mathbf{b} - \mathbf{y}\|_2$. If the 4 distances are all below the desired threshold ε, then the hyperedge e is added to G. Note that by using this method to select edges, the user defined tolerance parameter ε is expressed in image units (e.g., pixels), which is arguably more intuitive than the parameter quantization scheme used in the generalized Hough transforms.

In order to obtain the set of feature correspondences from which to estimate the affine transformation we need to find the maximum clique in the auxiliary hypergraph G. However, in general a large maximal clique is enough. Clearly, due to the way in which G was constructed, the associations in the clique will all agree on the found affine transformation within an error of ε pixels, allowing for a very robust estimation of the parameters.

In Figure 4.15 we show a simplified auxiliary 4-graph generated from a very small set of SIFT features extracted from the two affinely distorted images. Each vertex represents a correspondence between features in the two images, i.e., node $A1$ represents a match between feature A of the first fish and feature 1 of the second. Note that, while correspondences $B2$, $D4$, $F6$, and $H8$ agree on the same transformation, the best group of coherent matches is represented by the set $A1$, $C3$,$D4$,$E5$, and $H8$.

4.3.3 Finding cliques in k-graphs

In 1965, Motzkin and Straus introduced a continuous characterization of cliques in graphs [117]. This result was generalized to hypergraphs in [34], where it is shown that maximal cliques of a k-graph G are in one-to-one correspondence with the local solutions of the following program:

$$\min_{\mathbf{x}\in\Delta} L_{\bar{G}}(\mathbf{x}) + \frac{1}{k(k-1)} \sum_{i=1}^{n} x_i^k \tag{4.27}$$

where $\Delta = \{\mathbf{x} \in \mathbb{R}^n : \sum_{i=1}^n x_i = 1,\ \mathbf{x} \geq \mathbf{0}\}$ is the *standard simplex* and $L_{\bar{G}}(\mathbf{x}) = \sum_{e\in\bar{E}} \prod_{i\in e} x_i$ is the *Lagrangian* of $\bar{G}$. More precisely, $\mathbf{x}$ is a local/global solution of (4.27) if and only if it is the characteristic vector of a maximal/maximum clique C of G, i.e. $x_i = |C|^{-1}$ for all $i \in C$ and 0 elsewhere.

This result permits to cast clique problems on k-graphs in a continuous optimization setting. To solve Program 4.27 we turn it into an equivalent maximization of a homogeneous polynomial P with nonnegative coefficients over the standard simplex, where

$$P(\mathbf{x}) = \frac{1}{k(k-1)} \left[\left(\sum_{i=1}^{n} x_i \right)^k - \sum_{i=1}^{n} x_i^k \right] - L_{\bar{G}}(\mathbf{x}).$$

The function is then maximized using the discrete dynamics $x_j \leftarrow \alpha x_j \partial_j P(\mathbf{x})$, which, by means of the Baum-Eagon theorem [24], can be shown to be a growth transformation for P in Δ. Here, ∂_j denotes partial derivative with respect to x_j and α is the normalizing constant that projects $\mathbf{x}$ on Δ. By unfolding the partial derivative, we obtain

$$x_j \leftarrow \alpha x_j \left[\frac{1 - x_j^{k-1}}{k-1} - \partial_j L_{\bar{G}}(\mathbf{x}) \right]. \tag{4.28}$$

Since in our experiments $|E| \ll |\bar{E}|$, the computation of $\partial_j L_{\bar{G}}(\mathbf{x})$ is expensive, however, we can express this partial derivative in terms of G using the fact that $\partial_j L_{\bar{G}}(\mathbf{x}) = \partial_j L_{\mathcal{K}_n^k}(\mathbf{x}) - \partial_j L_G(\mathbf{x})$. Restricting our attention to 4-graphs, we have

$$\partial_j L_{\mathcal{K}_n^4}(\mathbf{x}) = \frac{(1-x_j)^3}{6} - \frac{1-x_j}{2} \sum_{\substack{i=1 \\ i\neq j}}^{n} x_i^2 - \frac{1}{3} \sum_{\substack{i=1 \\ i\neq j}}^{n} x_i^3.$$

Since the stable stationary points of the discrete dynamics (4.28) are in one-to-one correspondence with characteristic vectors of maximal cliques of G, they can be used as heuristic for the maximum clique problem on k-graphs, giving an approach with complexity $O(m(|V|+|E|))$ where m is the average number of iteration required by the dynamics for convergence.

4.3.4 Experimental results

In order to assess the precision improvement of our approach over current techniques, we compared our algorithm with the Hough transform approach proposed by Lowe in [102] for clustering pose votes. The experiments were conducted on the fishes database used in [90].

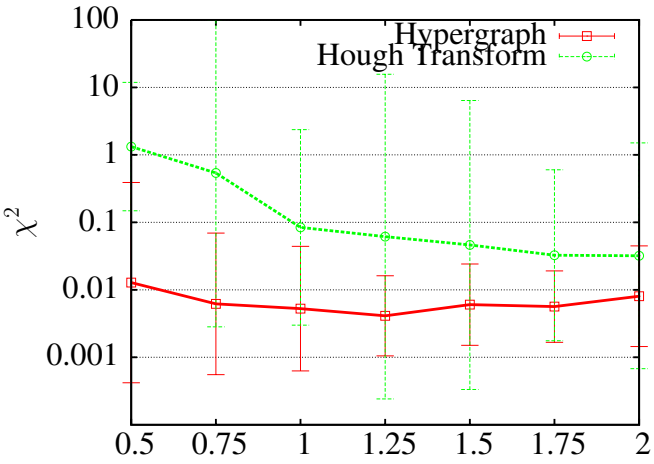
Since Lowe's method is more robust with deformations that preserve the aspect ratio, we evaluated the approaches on both aspect ratio preserving as well as aspect ratio deforming transformations. We run the algorithms on 20 randomly selected images under affine transformations at varying scales, rotations, aspect ratios and translations. We used the SIFT algorithm to extract features both from the original as well as the deformed images. The set of feasible associations, and hence, the vertex set of the auxiliary hypergraph, was generated by keeping the 100 best associations ranked according to their Euclidean distance in the feature space. For Lowe's algorithm we used the modified version of the best-bin-first algorithm as suggested in his paper.

We performed tests on transformations with scaling factors varying from 0.5 to 2, rotation angles from 0.5 to 3 radians, aspect ratios from 0.8 to 1.3 and random translations. Further, to assess the effect of the tolerance ϵ on our approach, we iterated the set of experiments with values of ϵ varying from 1 to 5 pixels. The quality of the affine transformations were evaluated by calculating the χ^2 distance between the true parameters and the estimated ones.

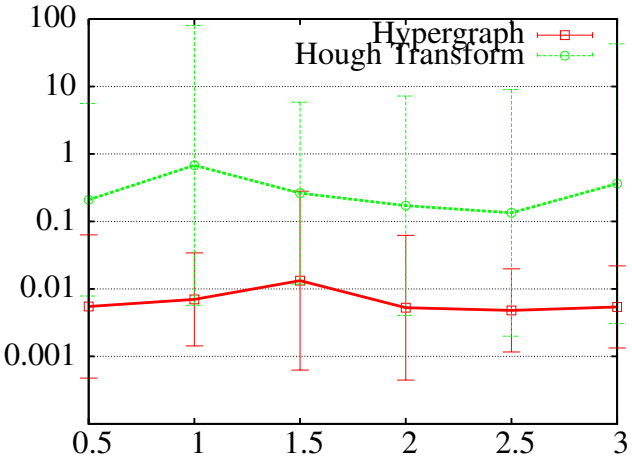
In Figures 4.16 the performance of the two approaches are compared. Figures 4.16(a) and 4.16(b) refer to aspect ratio preserving experiments, and plot the average error at varying scale and rotation respectively. These were calculated averaging over all the values of the other aspect preserving parameters and all the images. Conversely, Figure 4.16(c) presents the error at varying aspect ratio, averaging this time over all parameters and images. The performance of our approach is roughly one order of magnitude better than that obtained using the Hough transform. This is mainly because our hypergraph-based formulation does not need to define bins, or to select good initial matches, as it is able to capture the largest set of coherent associations in a more accurate way.

In Figure 4.17 we show how the performance of our approach varies with respect to the parameter ϵ. We can see that the algorithm is robust with respect to this parameter, in fact the accuracy obtained by using a tolerance of 3, 4 or 5 pixels does not change much and even with very relaxed constraints it is able to find highly coherent cliques. By the converse, setting very small values of ϵ is not only unnecessary, but also counterproductive: By setting a transformation constraint of 2 pixels the algorithm begins to reduce the number of hyperedges in G, yielding smaller cliques and, thus a less reliable estimation of the transformation parameters.

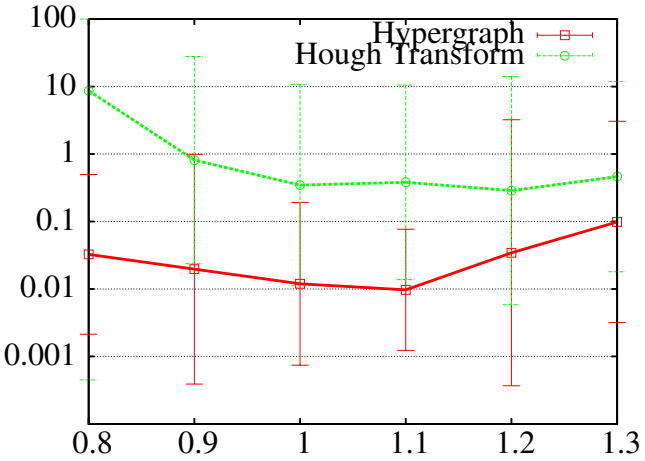
To illustrate how the differences in the χ^2 measure impact the estimation of the affine transformation, we show some examples in Table 4.3. Each row presents a different example, the first column shows the original fish, the second displays the target fish (black) and the original fish distorted with the transformation estimated using Lowe's approach (yellow). Finally, the last column shows the target fish (black) and the original fish dis-



(a) Scale



(b) Rotation



(c) Aspect Ratio

Figure 4.16: Performace of the algorithms

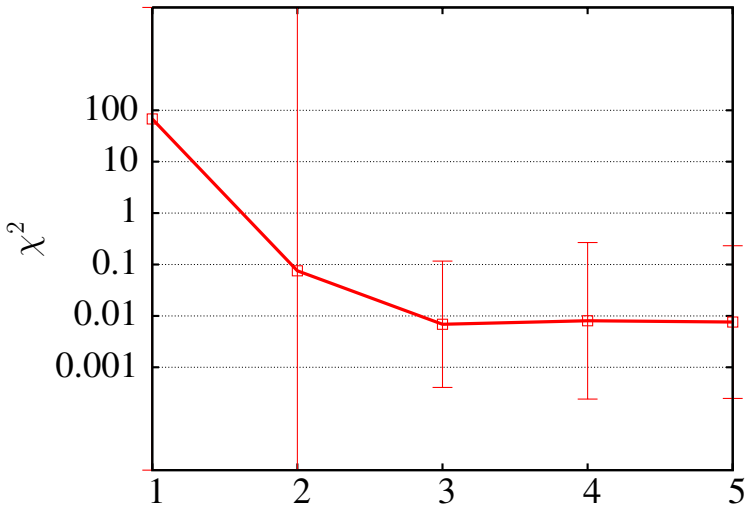


Figure 4.17: Sensitivity to ϵ

torted with the transformation estimated using our approach (yellow). From the images we can see that the estimation obtained by our method is indistinguishable from the target image, while the performances obtained using the Hough transform deviates from the target images significantly. Note that in the last two rows the same fish was transformed using respectively an aspect-ratio invariant transformation and an affine transformation that modified the aspect-ratio. As mentioned before, the solution proposed by Lowe is less robust with the latter transformation and, in fact, the error in the estimation of the last transformation is fairly high, while our approach performs very well in both cases.

4.3.5 Concluding Remarks

In this section we presented a new feature-based approach for the estimation of the transformation parameters between affinely distorted images. This is done by casting the problem to the search for a maximum clique in an auxiliary hypergraph, and then turning this into a continuous optimization problem over the standard simplex. Finally, we compared our approach with a state-of-the-art algorithm on a database of affinely distorted fish images. The tests show that our approach outperforms the competition in terms of precision of the estimation.

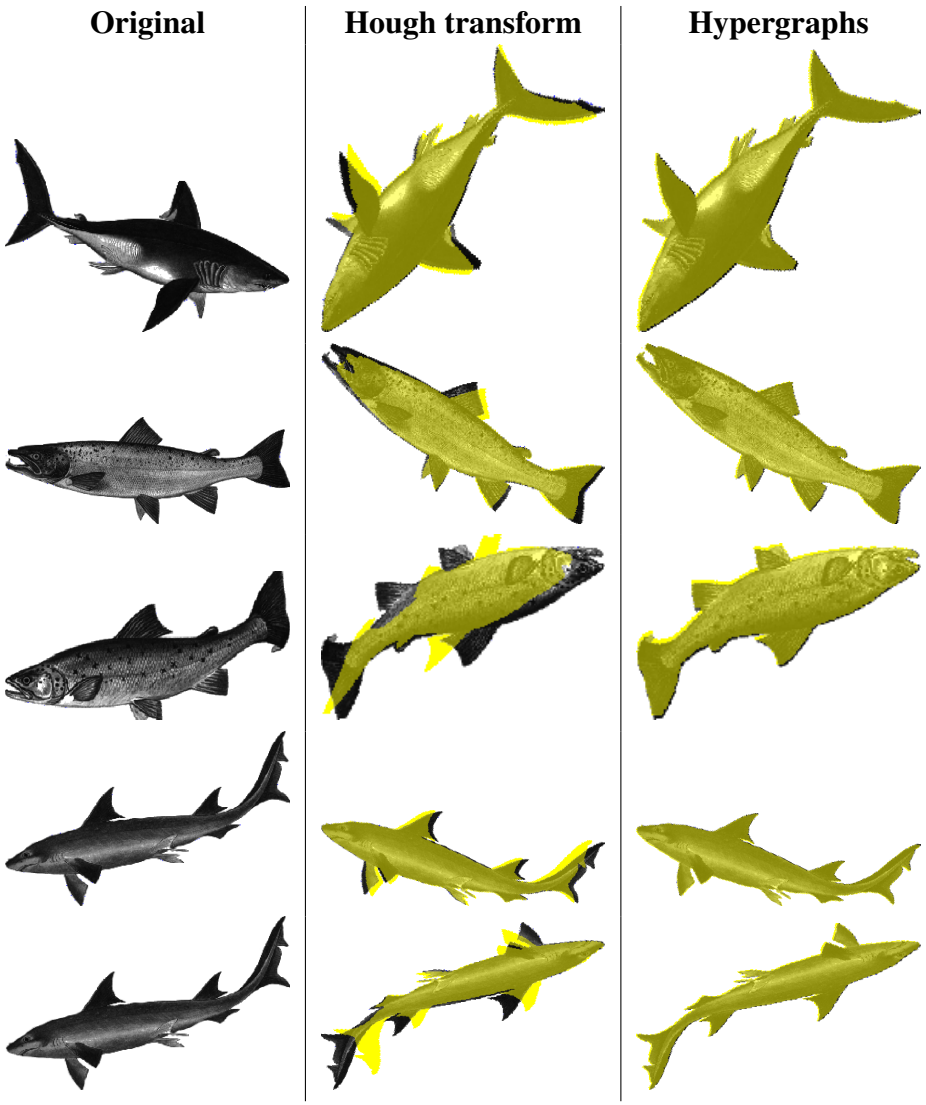


Table 4.3: Some example results.

5

Applications to 3D Surface Processing

In this chapter we explore the applications of our framework in the field of the processing of 3D data coming from structured light scanners. Specifically, in section 5.1 we present a novel technique for simultaneous coarse and fine surface registration. In section 5.2 an inlier selection scheme based on a special purpose association graph is adopted in order to find the best symmetry plane in a point cloud. Finally section 5.3 is dedicated to a novel phase shift pattern strategy that allows to obtain an unambiguous pixel coding with a minimal number of structured light projections.

5.1 Fine Surface Registration without Initial Motion Estimation

Surface registration is a fundamental step in the reconstruction of three-dimensional objects. This is typically a two step process where an initial coarse motion estimation is followed by a refinement. Most coarse registration algorithms exploit some local point descriptor that is intrinsic to the shape and does not depend on the relative position of the surfaces. By contrast, refinement techniques iteratively minimize a distance function measured between pairs of selected neighboring points and are thus strongly dependent on initial alignment. In this section we propose a novel technique that allows to obtain a fine surface registration in a single step, without the need of an initial motion estimation. The main idea of our approach is to cast the selection of correspondences between points on the surfaces in a game-theoretic framework, where a natural selection process allows mating points that satisfy a mutual rigidity constraint to thrive, eliminating all the other correspondences. This process yields a very robust inlier selection scheme that does not depend on any particular technique for selecting the initial strategies as it relies only on the global geometric compatibility between correspondences. The practical effectiveness of the proposed approach is confirmed by an extensive set of experiments and comparisons with state-of-the-art techniques.

5.1.1 Coarse and Fine Registration Techniques

The distinction between coarse and fine surface registration techniques is mainly related to the different strategies adopted to find pairs of mating points to be used for the estimation of the rigid transformation. Almost invariably, fine registration algorithms exploit an initial guess in order to constrain the search area for compatible mates and minimize the risk of selecting outliers. On the other hand, coarse techniques, which cannot rely on any motion estimation, must adopt a mating strategy based on the similarity between surface-point descriptors or resort to random selection schemes. The tension between the precision required for fine alignment versus the recall needed for an initial motion estimation stands as the main hurdle to the unification of such approaches.

The large majority of currently used fine alignment methods are modifications to the original ICP proposed by Zhang [189] and Besl and McKay [26]. These variants generally differ in the strategies used to sample points from the surfaces, reject incompatible pairs, or measure error. In general, the precision and convergence speed of these techniques is highly data-dependent and very sensitive to the fine-tuning of the model parameters. Several approaches that combine these variants have been proposed in the literature in order to overcome these limitations (see [136] for a comparative review). Some recent variants avoid hard culling by assigning a probability to each candidate pair by mean of evolutionary techniques [101] or Expectation Maximization [70]. ICP variants, being iterative algorithms based on local, step-by-step decisions, are very susceptible to the presence of local minima. Other fine registration methods include the well-known method by Chen [41] and signed distance fields matching [111].

Coarse registration techniques can be roughly classified in methods that exploit some global property of the surface, such as PCA [48] or Algebraic Surface Model [156], and methods that use some 3D feature descriptor to find plausible candidates pairs over the model and data surfaces. Global techniques are generally very sensitive to occlusion. Feature-based approaches are more precise and can align surfaces that exhibit only partial overlap. Nevertheless, the unavoidable localization error of the feature points prevents them from obtaining accuracies on par with fine registration methods. Among the most successful descriptors are Point Signatures [44] and Spin Images [89]. A completely different coarse registration approach is the RANSAC-based DARCES [39], which is based on the random extraction of sets of mates from the surfaces and their validation based on the accuracy of the estimated transformation. A recent and extensive review of all the different methods can be found in [138].

Regardless of the criteria used to obtain pairs of mating points, the subsequent step in surface registration is to search for the rigid transformation that minimizes the squared distance between them. Since many mature techniques are available to do this (for instance [80]), Our effort is toward the matching step itself: specifically by proposing a novel game-theoretic approach that is able to deal equally well with both coarse and fine registration scenarios.

5.1.2 Game-Theoretic Surface Registration

We are looking for a robust set of inliers for correspondence selection from which we can estimate the rigid transformation. Most of the currently adopted matching schemes operate on a local level, and global information comes only as an afterthought by checking the quality of the candidate matches with respect to the registration error obtained. The approach we are proposing, on the other hand, brings global information into the matching process by favoring sets of point-associations that are mutually compatible with a single rigid transformation. Fundamental to our approach is the fact that requiring the compatibility to a single transformation is equivalent to requiring that there exists a compatible transformation for each pair of mates. Following [161, 12], we model the mating process in a game-theoretic framework, where two players extracted from a large population select a pair of corresponding points from two surfaces to be registered with one another. The player then receives a payoff from the other players proportional to how compatible his pairings are with respect to the other player's choice, where the compatibility derives from the existence of a common rigid transformation. More explicitly, if there exists a rigid transformation that moves both his point and the other player's point close to the corresponding mates, then both players receive a high payoff, otherwise the payoff will be low. Clearly, it is in each player's interest to pick correspondences that are compatible with the mates the other players are likely to choose. In general, as the game is repeated, players will adapt their behavior to prefer matings that yield larger payoffs, driving all inconsistent hypotheses to extinction, and settling for an equilibrium where the pool of mates from which the players are still actively selecting their associations forms a cohesive set with high mutual support. Within this formulation, the solutions of the matching problem correspond to evolutionary stable states (ESS's), a robust population-based generalization of the notion of a Nash equilibrium.

In a sense, this mating process can be seen as a contextual voting system, where each time the game is repeated the previous selections of the other players affect the future vote of each player in an attempt to reach consensus. This way the evolving context brings global information into the selection process.

5.1.3 Non-cooperative Games

Originated in the early 40's, Game Theory was an attempt to formalize a system characterized by the actions of entities with competing objectives, which is thus hard to characterize with a single objective function [173]. According to this view, the emphasis shifts from the search of a local optimum to the definition of equilibria between opposing forces. In this setting multiple players have at their disposal a set of strategies and their goal is to maximize a payoff that depends also on the strategies adopted by other players. Evolutionary game theory originated in the early 70's as an attempt to apply the principles and tools of game theory to biological contexts.

Evolutionary game theory considers an idealized scenario where pairs of individuals

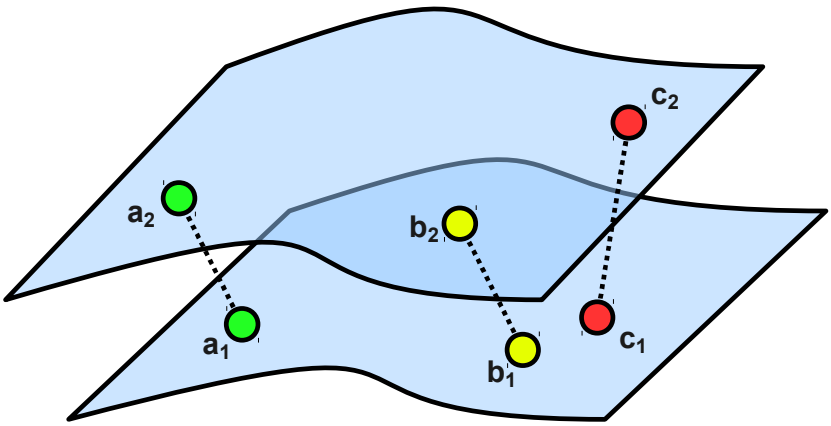


Figure 5.1: Example of mating strategies

are repeatedly drawn at random from a large population to play a two-player game. In contrast to traditional game-theoretic models, players are not supposed to behave rationally, but rather they act according to a pre-programmed behavior, or mixed strategy. It is supposed that some selection process operates over time on the distribution of behaviors favoring players that receive higher payoffs.

More formally, let $O = \{1, \cdots, n\}$ be the set of available strategies (*pure strategies* in the language of game theory) and, $C = (c_{ij})$ be a matrix specifying the payoff that an individual playing strategy i receives against someone playing strategy j. A *mixed strategy* is a probability distribution $\mathbf{x} = (x_1, \ldots, x_n)^T$ over the available strategies O. Clearly, mixed strategies are constrained to lie in the n-dimensional standard simplex

$$\Delta^n = \left\{ \mathbf{x} \in \mathbb{R}^n \ : \ x_i \geq 0 \text{ for all } i \in 1 \ldots n, \ \sum_{i=1}^{n} x_i = 1 \right\} .$$

The *support* of a mixed strategy $\mathbf{x} \in \Delta$, denoted by $\sigma(\mathbf{x})$, is defined as the set of elements chosen with non-zero probability: $\sigma(\mathbf{x}) = \{i \in O \mid x_i > 0\}$. The expected payoff received by a player choosing element i when playing against a player adopting a mixed strategy $\mathbf{x}$ is $(C\mathbf{x})_i = \sum_j c_{ij} x_j$, hence the expected payoff received by adopting the mixed strategy $\mathbf{y}$ against $\mathbf{x}$ is $\mathbf{y}^T C\mathbf{x}$. The *best replies* against mixed strategy $\mathbf{x}$ is the set of mixed strategies

$$\beta(\mathbf{x}) = \{\mathbf{y} \in \Delta \mid \mathbf{y}^T C\mathbf{x} = \max_{\mathbf{z}} (\mathbf{z}^T C\mathbf{x})\} .$$

A strategy $\mathbf{x}$ is said to be a *Nash equilibrium* if it is the best reply to itself, i.e., $\forall \mathbf{y} \in \Delta, \ \mathbf{x}^T C\mathbf{x} \geq \mathbf{y}^T C\mathbf{x}$. This implies that $\forall i \in \sigma(\mathbf{x})$ we have $(C\mathbf{x})_i = \mathbf{x}^T C\mathbf{x}$; that is, the payoff of every strategy in the support of $\mathbf{x}$ is constant.

A strategy $\mathbf{x}$ is said to be an *evolutionary stable strategy* (ESS) if it is a Nash equilibrium and

$$\forall \mathbf{y} \in \Delta \quad \mathbf{x}^T C\mathbf{x} = \mathbf{y}^T C\mathbf{x} \Rightarrow \mathbf{x}^T C\mathbf{y} > \mathbf{y}^T C\mathbf{y} . \tag{5.1}$$

This condition guarantees that any deviation from the stable strategies does not pay.

5.1.4 Mating Strategies and Payoffs

Central to this framework is the definition of a *mating game*, which implies the definition of the strategies available to the players and of the payoffs related to these strategies. Given a set of model points M and a set of data points D we call a *mating strategy* any pair (a_1, a_2) with $a_1 \in M$ and $a_2 \in D$. We call the set of all the mating strategies S. In principle, all the model and data points could be used to build the mating strategies set, thus giving $S = M \times D$. In practice, however, we adopt some heuristics that allow us to obtain good alignments with a much smaller set. Once S has been selected, our goal becomes to extract from it the largest subset that includes only correctly matched points: that is, strategies that associate a point in the model surface with the same point in the data surface. To enforce this we assign to each pair of mating strategies a payoff that is inversely proportional to a measure of violation of the rigidity constraint. This violation can be expressed in several ways, but since all the rigid transformations preserve Euclidean distances, we choose this property to express the coherence between mating strategies.

Definition 1. *Given a function $\pi : S \times S \rightarrow \mathbb{R}^+$, we call it a* rigidity-enforcing payoff function *if for any $((a_1, a_2), (b_1, b_2))$ and $((c_1, c_2), (d_1, d_2)) \in S \times S$ we have that $||a_1 - b_1| - |a_2 - b_2|| > ||c_1 - d_1| - |c_2 - d_2||$ implies $\pi((a_1, a_2), (b_1, b_2)) < \pi((c_1, c_2), (d_1, d_2))$. In addition, if $\pi((a_1, a_2), (b_1, b_2)) = \pi((b_1, b_2), (a_1, a_2))$, π is said to be* symmetric.

A rigidity-enforcing payoff function is a function that is monotonically decreasing with the absolute difference of the Euclidean distances between respectively the model and data points of the mating strategies compared. In other words, given two mating strategies, their payoff should be high if the distance between the model points is equal to the distance between the data points and it should decrease as the difference between such distances increases. In the example of Figure 5.1 mating strategies (a_1, a_2) and (b_1, b_2) are coherent with respect to the rigidity constraint, whereas (b_1, b_2) and (c_1, c_2) are not, thus it is expected that $\pi((a_1, a_2), (b_1, b_2)) > \pi((b_1, b_2), (c_1, c_2))$.

Further, if we want mating to be one-to-one, we must put an additional constraint on the payoffs, namely that mates sharing a point are incompatible.

Definition 2. *A rigidity-enforcing payoff function π is said to be* one-to-one *if $a_1 = b_1$ or $a_2 = b_2$ implies $\pi((a_1, a_2), (b_1, b_2)) = 0$.*

Given a set of mating strategies S and an enumeration $O = \{1, ..., |S|\}$ over it, a *mating game* is a non-cooperative game where the population is defined as a vector $\mathbf{x} \in \Delta^{|S|}$ and the payoff matrix $C = (c_{ij})$ is defined as $c_{ij} = \pi(s_i, s_j)$, where $s_i, s_j \in S$ are enumerated by O and π is a symmetric one-to-one rigidity-enforcing payoff function. Intuitively, $\mathbf{x}_i$ accounts for the percentage of the population that plays the i-th mating strategy. By using a symmetric one-to-one payoff function in a mating game we are guaranteed that ESS's will not include mates sharing either model or data nodes. In fact, given a non-negative payoff function, a stable state cannot have in its support a pairs of strategies with payoff 0 [12]. Moreover, a mating game exhibits some additional interesting properties.

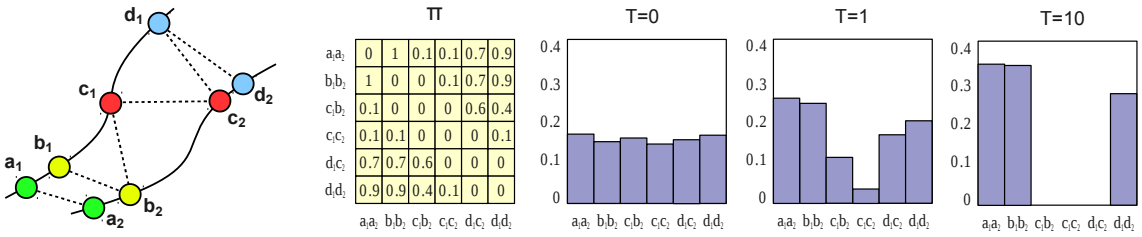


Figure 5.2: An example of the evolutionary process. Four points are sampled from the two surfaces and a total of six mating strategies are selected as initial hypotheses. The matrix Π shows the compatibilities between pairs of mating strategies according to a one-to-one rigidity-enforcing payoff function. Each mating strategy got zero payoff with itself and with strategies that share the same source or destination point (i.e., $\pi((b_1, b_2), (c_1, b_2)) = 0$). Strategies that are coherent with respect to rigid transformation exhibit high payoff values (i.e., $\pi((a_1, a_2), (b_1, b_2)) = 1$ and $\pi((a_1, a_2), (d_1, d_2)) = 0.9$)), while less compatible pairs get lower scores (i.e., $\pi((a_1, a_2), (c_1, c_2)) = 0.1$). Initially (at T=0) the population is set to the barycenter of the simplex and slightly perturbed. After just one iteration, (c_1, b_2) and (c_1, c_2) have lost a significant amount of support, while (d_1, c_2) and (d_1, d_2) are still played by a sizable amount of population. After ten iterations (T=10) (d_1, d_2) has finally prevailed over (d_1, c_2) (note that the two are mutually exclusive). Note that in the final population $((a_1, a_2), (b_1, b_2))$ have a higher support than (d_1, d_2) since they are a little more coherent with respect to rigidity.

Theorem 5. *Given a set of model points M, a set of data points $D = TM$ that are exact rigid transformations of the points in M, and a set of mating strategies $S \subseteq M \times D$ with $(m, Tm) \in S$ for all $m \in M$, and a mating game over them with a payoff function π, the vector $\hat{x} \in \Delta^{|S|}$ defined as*

$$\hat{x}_i = \begin{cases} 1/|M| & \text{if } s_i = (m, Tm) \text{ for some } m \in M\,; \\ 0 & \text{otherwise,} \end{cases}$$

is an ESS and obtains the global maximum average payoff.

Sketch of proof. Let $\hat{S} \subseteq S$ be the set of mates that match a point to its copy, clearly for all $s, q \in \hat{S}, s \neq q$ we have $\pi(s, q) = 1$, while for $s \in \hat{S}$ and $q \in S \setminus \hat{S}$, we have $\pi(s, q) < 1$. For all $s \in \hat{S}$ we have that $\pi(\hat{x}, \hat{x}) = \frac{|M|-1}{|M|}$ while, since π is one-to-one, for any $q \in S \setminus \hat{S}$ there must be at least one $s_q \in \hat{S}$ with $\pi(q, s_q) = 0$, thus $\pi(q, \hat{x}) < \frac{|M|-1}{|M|}$, thus $\hat{x}$ is a Nash equilibrium. Further, since the inequality is strict, it is an ESS. Finally, $\hat{x}$ is a global maximizer of π since $\frac{|M|-1}{|M|}$ is the maximum value that a one-to-one normalized payoff function over $|M|$ points can attain. □

This theorem states that when matching a surface with a rigidly transformed copy of itself the optimal solution (i.e., the population configuration that selects all the mating strategies assigning each point to its copy) is the stable state of maximum payoff. Since well established algorithms to evolve a population to such a state exist, this provides us

with an effective mating approach. Clearly, aligning a surface to an identical copy is not very useful in practical scenarios, where occlusion and measurement noise come into play. While the quality of the solution in presence of noise will be assessed experimentally, we can give some theoretical results regarding occlusions.

Theorem 6. *Let M be a set of points with $M_a \subseteq M$ and $D = TM_b$ a rigid transformation of $M_b \subseteq M$ such that $|M_a \cap M_b| \geq 3$, and $S \subseteq M_a \times D$ be a set of mating strategies over M_a and D with $(m, Tm) \in S$ for all $m \in M_a \cap M_b$. Further, assume that the points that are not in the overlap, that is the points in $E_a = M_a \backslash (M_a \cap M_b)$ and $E_b = M_b \backslash (M_a \cap M_b)$, are sufficiently far away such that for every $s \in S, s = (m, Tm)$ with $m \in M_a \cap M_b$ and every $q \in S, q = (m_a, Tm_b)$ with $m_a \in E_a$ and $m_b \in E_b$, we have $\pi(q, s) < \frac{|M_a \cap M_b|-1}{|M_a \cap M_b|}$, then, the vector $\hat{x} \in \Delta^{|S|}$ defined as*

$$\hat{x}_i = \begin{cases} 1/|M| & \text{if } s_i = (m, Tm) \text{ for some } m \in M_a \cap M_b\,; \\ 0 & \text{otherwise,} \end{cases}$$

is an ESS.

Sketch of proof. We have $\pi(\hat{x}, \hat{x}) = \frac{|M_a \cap M_b|-1}{|M_a \cap M_b|}$. Let $q \in S$ be a strategy not in the support of $\hat{x}$, then, either it maps a point in M_a or M_b, thus receiving payoff $\pi(q, \hat{x}) < \frac{|M_a \cap M_b|-1}{|M_a \cap M_b|}$ because of the one-to-one condition, or it maps a point in E_a to a point in E_b, receiving, by hypothesis, a payoff $\pi(q, \hat{x}) < \frac{|M_a \cap M_b|-1}{|M_a \cap M_b|}$. Hence, $\hat{x}$ is an ESS. □

The result of theorem 6 is slightly weaker than theorem 5, as the face of the simplex corresponding to the "correct" overlap, while being an evolutionary stable state, is not guaranteed to obtain the overall highest average payoff. This is not a limitation of the framework as this weakening is actually due to the very nature of the alignment problem itself. The inability to guarantee the maximality of the average payoff is due to the fact that the original object (M) could contain large areas outside the overlapping subset that are perfectly identical. Further, objects that are able to slide (for instance a plane or a sphere) could allow to move between different mixed strategies without penalty. These situations cannot be addressed by any algorithm without relying on supplementary information. However, in practice, they are quite unlikely extreme cases. In the experimental section we will show that our approach can effectively register even quasi-planar surfaces.

5.1.5 Building the Mating Strategies Set

From a theoretical point of view the total number of mating strategies in a registration problem is $|M \times D|$, which can be very large even with medium-sized surfaces. In practice it is possible to apply several heuristics to select a lower number of candidates while still achieving good alignment results. Since the proposed approach is very selective it is not necessary to use all the model points: even a highly aggressive subsampling does not affect the registration quality, provided that some points in the overlapping region between model and data are retained. In fact, our approach does not try to find a good registration by means of a vote counting validation; instead it takes quite the opposite route, by

self-validating the selected mixed strategy exploiting its internal coherence. Once the model points have been subsampled, the mating strategies set could be created by pairing each one of them with all the data points. Again, while this approach would work, it is somewhat wasteful since most of the mating strategies could be dropped on the basis of some local property of the surface surrounding the model and data point. For instance, the mean or Gaussian curvatures can be compared or some surface feature can be calculated in order to select only meaningful pairs. In the experimental section we will suggest an effective selection strategy.

Once a proper set of mates has been chosen, a payoff function is needed. In principle, any proper one-to-one symmetric rigidity-enforcing payoff function could be used to capture the coherence between pairs of mating strategies. From a practical point of view it is often advisable to use bounded functions, usually in the interval $]0, 1]$. Very good candidates are negative exponentiations of the difference between the distances of the model and data points or the ratio between the min and the max distance. In general, the steeper is the function, the more selective is the choice of the inlier mating strategies.

5.1.6 Evolving to an Optimal Solution

The search for a stable state is performed by simulating the evolution of a natural selection process. Under very loose conditions, any dynamics that respect the payoffs is guaranteed to converge to Nash equilibria [173] and (hopefully) to ESS's; for this reason, the choice of an actual selection process is not crucial and can be driven mostly by considerations of efficiency and simplicity. We chose to use the replicator dynamics, a well-known formalization of the selection process governed by the following equation

$$\mathbf{x}_i(t+1) = \mathbf{x}_i(t)\frac{(C\mathbf{x}(t))_i}{\mathbf{x}(t)^t C\mathbf{x}(t)} \tag{5.2}$$

where $\mathbf{x}_i$ is the i-th element of the population and C the payoff matrix. A simple but complete example of the evolution process is shown in Figure 5.2.

Once the population has reached a local maximum, all the non-extincted mating strategies can be used to calculate the rigid transformation between data and model surfaces. A clear advantage of our approach is that in the final mixed strategy each pair of points is weighted proportionally to its degree of participation in the equilibrium (see Figure 5.2). This is similar in spirit to the concept of compatibility between mates adopted by a number of fine registration algorithms, yet it does not depend at all on supplementary information such as surface normals or texture color. This compatibility can be used to weigh each pair when calculating the best surface alignment by using a weighted least squares fitting technique [80].

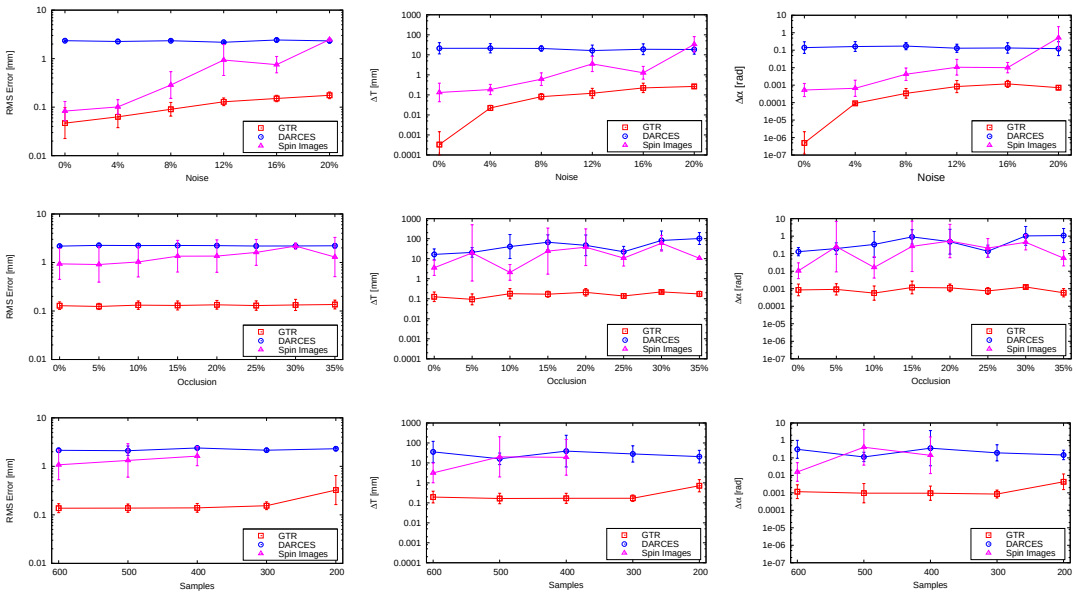


Figure 5.3: Comparison of coarse registration techniques using real range data.

5.1.7 Experimental Results

Since the proposed technique can be used independently for coarse and fine registration we evaluated its performance with respect to state-of-the-art algorithms of both fields. All the experiments have been executed on two sets of data: range images obtained from real-world scanners and synthetically-generated surfaces. For the first set of experiments we selected models from publicly available databases; specifically the Bunny [165], the Armadillo [96] and the Dragon [53] from the Stanford 3D scanning repository. To further assess the shortcomings of the various approaches, we used three synthetic surfaces representative of as many classes of objects: a wave surface, a fractal landscape and an incised plane (see Figure 5.4).

In all the experiments the set of mating strategies was obtained using the same selection technique. We used the MeshDOG [186] 3D feature detector to find interesting points in both the model and the data range images. A descriptor was associated to each point of interest; after considering both the MeshHOG and the Spin Image descriptors, we preferred the latter as we found it to be more distinctive. Then a set of candidate source points was subsampled from the model and for each source point we created 5 mating strategies by connecting it to the 5 points with the most compatible descriptors. The rigidity-enforcing payoff function chosen was

$$\pi((a_1, b_1), (a_2, b_2)) = \frac{min(|a_1 - a_2|, |b_1 - b_2|)}{max(|a_1 - a_2|, |b_1 - b_2|)} \tag{5.3}$$

where a_1, a_2, b_1 and b_2 are respectively the two model (source) and data (destination) points in the compared mating strategies.

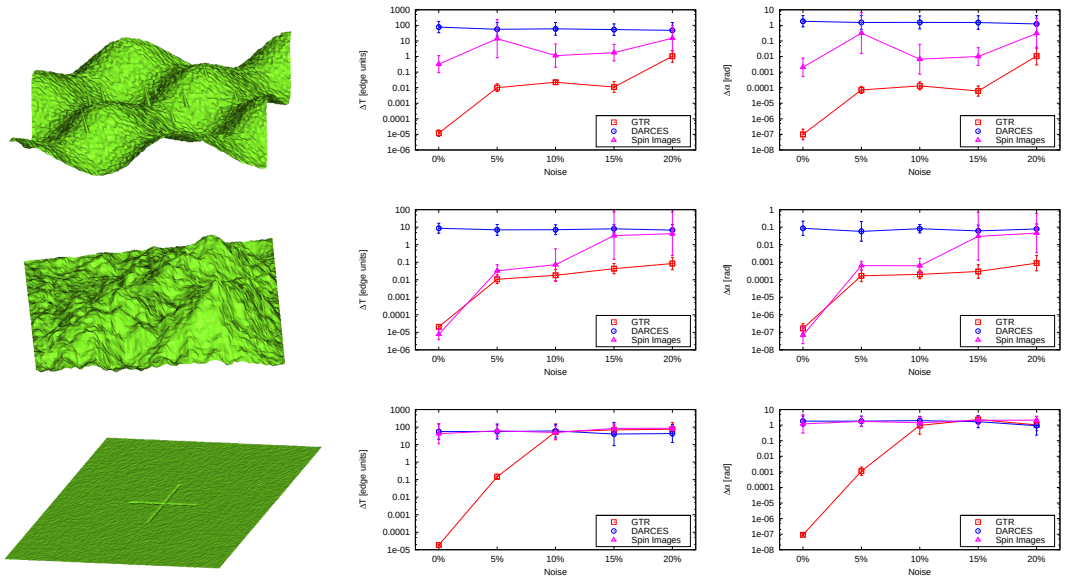


Figure 5.4: Comparison of coarse registration techniques using synthetic objects.

5.1.8 Coarse Registration

We compared our method with two widely used coarse registration methods: RANSAC-based DARCES [39] and Spin Images [89]. DARCES has been implemented according to the original paper, while we used the Spin Images variant suggested in [37] to obtain a higher accuracy. In Figure 5.3 we show the results obtained using the set of surfaces from the Stanford repository. Each test was made under different conditions of noise, occlusion and subsampling and was run for a total of 12 times over the set of range images. For each set of experiments we plot the RMS error and the estimation errors of the translation vector and rotation angle. In order to obtain a ground truth for precise error measurement we generated the data points by adding Gaussian noise, and random occlusion and motion to the model points. The surfaces used for these experiments were obtained from laser scans of objects of the size of hundreds of millimeters with a resolution of about one tenth of millimeter.

The first row of Figure 5.3 plots the sensitivity to Gaussian noise exhibited by the different techniques. The noise level is expressed as the ratio between the standard deviation of the noise and the average edge length. While DARCES is not very sensitive to noise, it delivers by far the worst overall results. By contrast, Spin Images give fairly good results at low noise levels, but their performance worsens quickly as noise is increased. The proposed approach, on the other hand, exhibits errors that are consistently an order of magnitude below Spin Images. In the second row of plots we show the effect of occlusion. This set of experiments was performed with a constant level of Gaussian noise with standard deviation equal to 12% of the average edge length. The results show that the tested techniques are substantially insensitive to occlusions, our technique constantly outperforming the other approaches. Finally, the third row shows the effect of subsam-

pling. Our game-theoretical registration method outperforms the other approaches. Note that the Spin Images based technique was never able to find a correct transformation when provided with less than 300 samples.

Figure 5.4 plots the alignment results on the three synthetic surfaces. Each set of experiments was conducted over a single type of surface (displayed at the beginning of the row) with 12 runs for each technique and noise level. Since these objects are synthetic, errors on translation are expressed in edge units. The "wave" test object (first row) offers a regular surface with few outstanding features and high redundancy of the pattern; in this

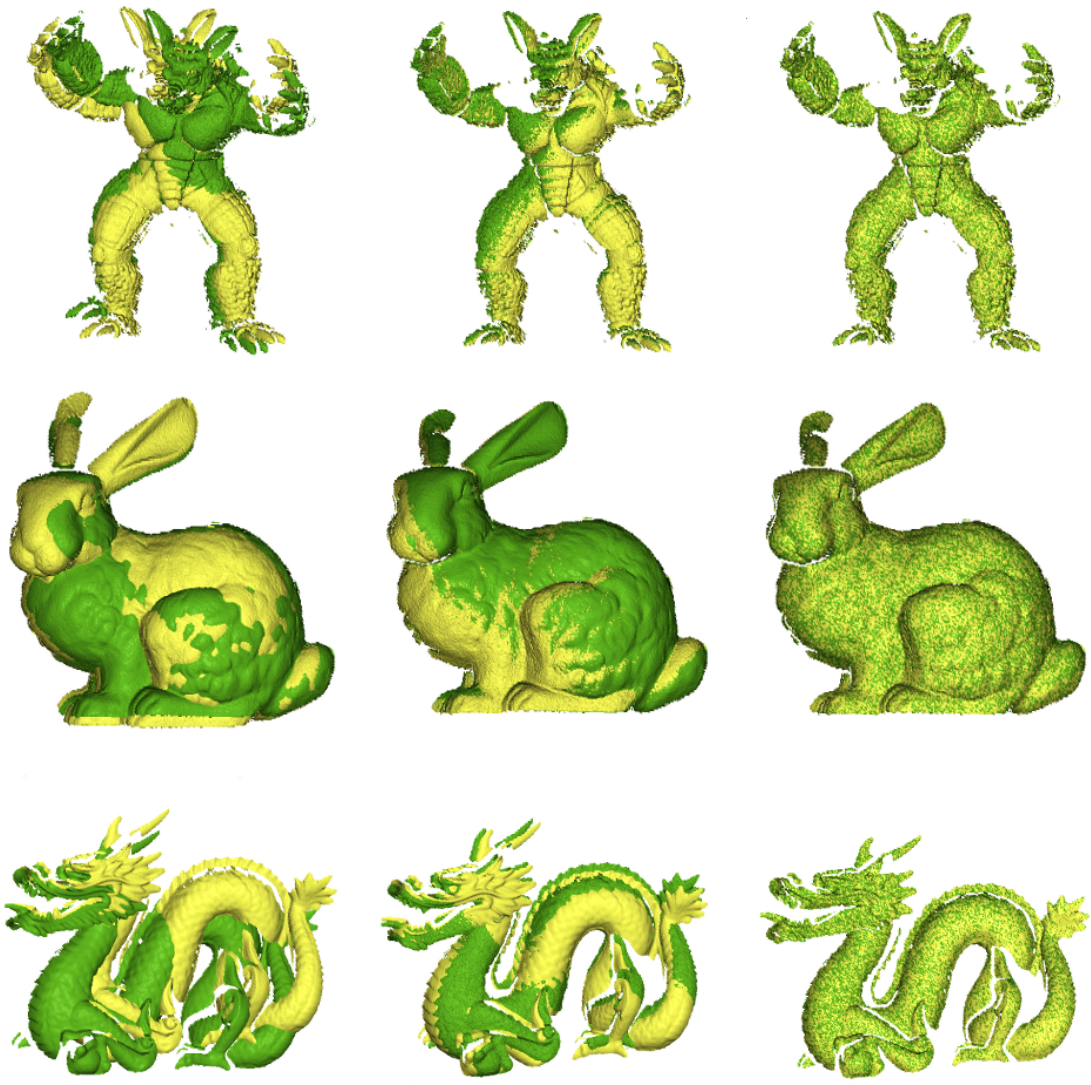


Figure 5.5: Examples of surface registration obtained respectively with RANSAC-based DARCES (first column), Spin Images (second column), and our game-theoretic registration technique (third column)

scenario the Spin Images technique is affected by the inability to discern among a large amount of similar descriptors, thus it performs poorly at all noise levels. Conversely, the geometric-based consensus exploited by our registration approach allows for a more precise selection and thus a more accurate registration. The "fractal landscape" test object (second row) is an irregular surface that allows to produce very distinctive feature descriptors. In fact with low levels of noise both Spin Images and our technique perform very well, albeit as noise increases we achieve better results. Finally, the "incised plane" object (third row) is a big flat domain with a small cross just half an edge deep. This represents a very difficult target for most registration techniques, since very few and faint features are available, while a large planar surface dominates the landscape. Despite the lack of good detectable points, our technique is able to register the surface as long as noise is minimal. With higher noise levels the bumped cross fades and becomes almost indistinguishable from the plane itself. Note that DARCES achieves mediocre results under all tested conditions.

5.1.9 Fine Registration

The performance of our approach with respect to fine registration has been studied in a separate batch of experiments. The goal of this test is two-fold: we want to evaluate our quality as a complete alignment tool and, at the same time, find the breaking point for traditional fine registration techniques. The method we used for comparison is a best-of-breed ICP variant, similar to the one proposed in [165]. Point selection is based on Normal Space Sampling [136], and point-surface normal shooting is adopted for finding correspondences; distant mates or candidates with back-facing normals are rejected. To minimize the influence of incorrect normal estimates, matings established on the boundary of the mesh are also removed. The resulting pairings are weighted with a coefficient based on compatibility of normals, and finally a 5%-trimming is used. Each test was performed by applying a random rotation and translation to different range images selected from the Stanford 3D scanning repository. Additionally, each range image was perturbed with a constant level of Gaussian noise with standard deviation equal to 12% of the average edge length. We completed 100 independent tests and for each of them we measured the initial RMS error between the ground-truth corresponding points and the resulting error after performing a full round of ICP (ICP) and a single run of our registration method (GTR). In addition we applied a step of ICP to the registration obtained with our method (GTR + ICP) in order to assess how much the solution extracted using our approach was further refinable. A scatter plot of the obtained errors before and after registration is shown in Figure 5.6. We observe that ICP reaches its breaking point quite early; in fact with an initial error above the threshold of about 20mm it is unable to find a correct registration. By contrast GTR is able to obtain excellent alignment regardless of the initial motion perturbation. Finally, applying ICP to GTR decreases the RMS only by a very small amount.

While we did not carry out any formal benchmark of the execution time required by our technique, we always observed a very fast convergence of the replicator dynamics,

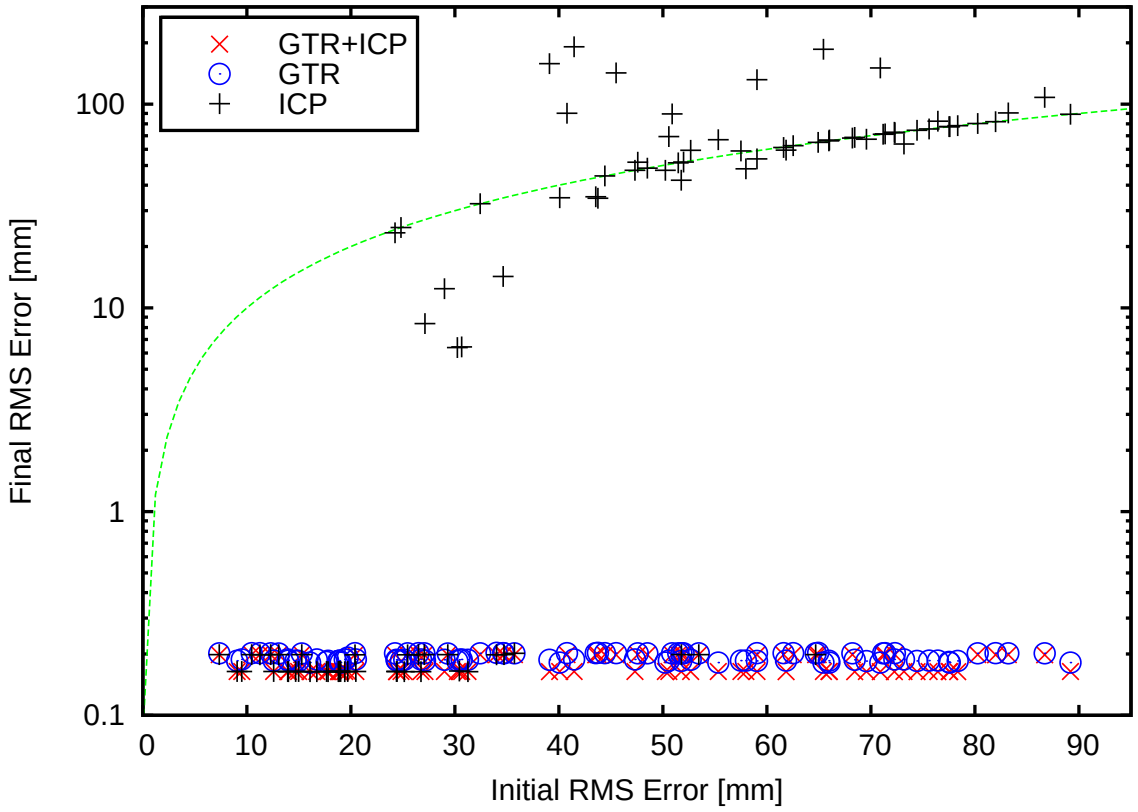


Figure 5.6: Comparison of fine registration accuracies (the green dashed line represents y=x). Graph best viewed in color.

even with several thousands of mating strategies. In the worst scenarios our unoptimized C++ implementation of the framework required less than 2 seconds (on a typical desktop PC) to evolve a population to a stable state.

5.1.10 Concluding Remarks

In this section we introduced a novel game-theoretic technique that solves both the coarse and fine surface registration problems at once. Our approach has several advantages over the state-of-the-art: it does not require any kind of initial motion estimation, as it does not rely on spatial relationships between model and data points, it does not need any threshold as it produces a continuous compatibility weight on the selected point matches that can be used directly for alignment estimation, and, differently from most inlier selection techniques, it is not affected by a high number of outliers since it operates an explicit search for good inliers rather than using random selection or vote counting for validation. From a theoretical point of view, a sound correspondence between optimal alignments and evolutionary equilibria has been presented and a wide range of experiments validated both

the robustness of the approach with respect to noise and its performance in comparison with other well-known techniques.

5.2 Consensus Graph for Symmetry Plane Estimation

Finding the most relevant symmetry planes for an object is a key step in many computer vision and object recognition tasks. In fact such information can be effectively used as a starting point for object segmentation, noise reduction, alignment and recognition. Some of these applications are strongly affected by the accuracy of symmetry planes estimation, thus the use of a technique that is both accurate and robust to noise is critical. In this section we introduce a new weighted association graph which relates the main symmetry planes of 3D objects to large sets of tightly coupled vertices. This technique allows us to cast symmetry detection to a classical pairwise clustering problem, which we solve using the very effective Dominant Sets framework. The improvement of our approach over other well known techniques is shown with several tests over both synthetic data and sampled point clouds.

5.2.1 Introduction

Among all the features that can be extracted from a 2D image or 3D object symmetry is one of the most important. For instance symmetry axes can be used to set a canonical orientation, which in turn can be useful for matching different objects, computing partial registration of surfaces or simply to offer a comparable point of view in visual application. Other applications exploit symmetries as a descriptor for searching in large databases or as a starting point for locating other features. In addition partial local symmetries can be very effective as a cue in segmentation. Finally, due to its abundance in both natural and crafted object, symmetries play an important role in visual perception and Gestalt theory. Because of this wide range of applications, symmetry estimation has been a popular research field for long time. Most seminal papers appeared in the eighties and were focused on perfect symmetries, mainly because this kind of analysis required less computational power. Atallah[19] and Wolter[177] solved the problem very efficiently by casting it into substring matching. Some more recent approaches use orientation histograms[151], singular value decomposition[142], extended Gaussian image[152] and generalized complex moments[144]. Also principal component analysis[104] can be used with good results when dealing with strongly symmetric objects. All these approaches rely on the existence of a perfect plane of symmetry, thus they are very sensitive to noise and cannot be applied to objects where partial or local symmetries are present.

In order to deal with partial or imperfect symmetries some measures of the distance between a given object and the nearest perfectly symmetric shape have been proposed[185, 98]. Such a measure could be used to search in the full plane space for the solution that minimizes the error, however since it can be applied to just one plane at a time it would not be an efficient solution.

A more practical approach, that has been proved effective in a wide range of contexts, is to use some kind of transform. That is a mapping from a function defined over the space of the object data to a function defined over the space of the feasible solutions, where local and global maximum are usually associated to local and global solutions of the problem itself. The most well known transform in computer vision is the Hough transform, which maps points on a plane to the space of parametrized straight lines. Over the last decade several symmetry transform have been proposed[133, 42, 28, 61, 103]. Recently the Planar-Reflective Symmetry Transform (PRST) has been presented in[129, 130] and extended in[137]. This is a transform that maps the points of three dimensional objects to a real-valued function in the three dimensional space of parametric planes. As for any transform technique in practical implementations it is computed in a discrete form: that is, the function is evaluated only on a large set of randomly selected points and the results are used to vote for the best solutions in a discrete set of bin that are uniformly distributed over the space of planes. This approach leads to coarse errors, due to the discrete size of the bins. However a continuous refinement technique is proposed that greatly enhances the solution by searching for a more accurate symmetry plane inside the winning bin by using an error minimization technique inspired by the Iterative Closest Point algorithm[26].

We propose a novel approach which casts the problem to the search for pairwise clusters in a newly defined weighted association graph that we named Symmetry Consensus Graph. In such a graph each vertex is associated to a candidate symmetry plane and the similarity between vertices expresses the consensus between planes. This way each large cluster in the graph captures a set of highly similar plane candidates and leaves out all the unwanted outliers, according to the high internal homogeneity and external inhomogeneity properties of clusters. In section 5.2.2 we define the Symmetry Consensus Graph in both a binary and a weighted form. In section 5.2.5 we suggest a very effective framework in order to find large cluster in a Consensus Graph. Finally, in sections 5.2.6 and 5.2.7 we compare our approach with PCA and the very accurate PRST technique.

5.2.2 The Symmetry Consensus Graph

Given a large set of 3D points describing an object our goal is to find the largest subsets that share a common symmetry plane. The most widely used approach is to select a large number of point pairs and thus collect votes for the plane between them in an appropriate parameters space. As for any other voting technique, concrete implementations of this approach need to choose a quantization of the parameters space into discrete bins. In many scenarios the choice of the size of the bins can be very critical as a too coarse quantization can lead to a poor precision in the estimation of the symmetry, while a too fine one could suffer of vote dispersion due to the dimensionality curse. In [130] the authors suggest to adopt a quite coarse parameters space quantization ad thus apply a continuous refinement step in order to minimize the local symmetry error among points that voted for the winning bin. This step enhances the quality of the solution found, but still is quite sensible to the presence of local outliers in the rather large winning bin. In

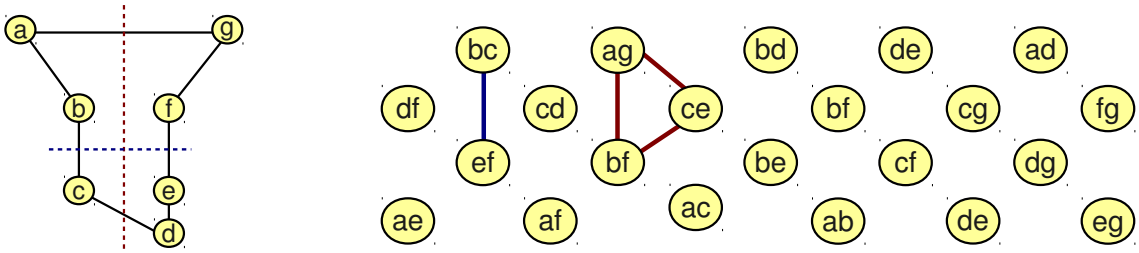


Figure 5.7: A very simple object with perfect partial symmetries (left) and its Binary Symmetry Consensus Graph (right). Only pairs of vertices that share the same symmetry plane are connected by an edge. For practical reasons only a part of the graph is shown.

order to obtain the best results an optimal bin size should be used, that is a quantization that allows to collect just the most relevant votes, leaving out local outliers. Unfortunately it is very hard to know in advance which size to use as it depends on many factors such as noise and quality of the symmetry. By contrast, the main idea which underpin our method is to avoid the quantization by using a more global approach able to weight all the voting point pairs at once.

5.2.3 Binary Symmetry Consensus Graph

In this section we focus our attention on objects that, although not being symmetric as a whole, contain one or more subset of points that are perfectly specular by a plane. Obviously this scenario is not very realistic from a practical point of view, nevertheless it permits us to define an useful structure that we will expand in section 5.2.4 to be more suitable for real world applications:

Definition 3. *Given a set of points in space* $P = \{p_1, p_2 \dots p_n\}$*, their Binary Symmetry Consensus Graph (BSCG) is the undirected graph* $G_B = (V_B, E_B)$ *where* $V_B = \{(p_a, p_b) \in P | a < b\}$ *is the set of pairs of points in* P *and where* $((p_a, p_b), (p_c, p_d)) \in E_B$ *if and only if the plane between* p_a *and* p_b *is exactly the same plane between* p_c *and* p_d*.*

Each vertex in such a graph represents a pair of points in the original object and is attributed with the plane that separates such points in a symmetric way. The key property of this graph is that two vertices are connected if and only if they share the same plane. As this is an equivalence relation the transitive property holds and it is easy to see that each set of vertices that share the same symmetry plane with each other forms a complete subgraph (clique). By building the BSCG over a set of points we can thus cast the search for the most relevant symmetry planes into the search for the largest cliques over the Consensus Graph. Although this is known to be a NP-complete problem many powerful heuristics have been suggested for it.

In figure 5.7 we illustrate a simple example of a BSCG. On the left an object with two partial perfect symmetry planes is shown. On the right the BSCG is built over all the pairs of point of the object. Only two cliques are found in the BSCG, respectively between pairs

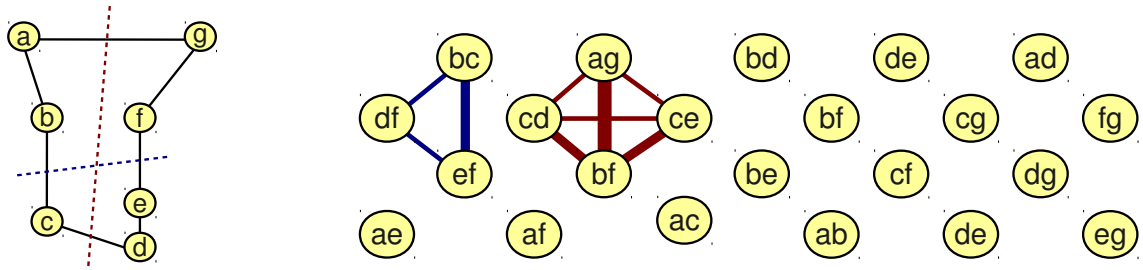


Figure 5.8: A very simple object without perfect symmetries (left) and its Weighted Symmetry Consensus Graph (right). The weight of the edges is proportional to their thickness. For practical reasons negligible edges are not drawn and only a part of the graph is shown.

ag, bf, ce, which are symmetric with respect to the vertical axis, and between pairs bc, ef, which are symmetric with respect to the horizontal axis. All other pairs do not share their symmetry plane with any other pair in the graph, thus all of them are singleton.

5.2.4 Weighted Symmetry Consensus Graph

The assumptions taken in section 5.2.3 are very strict in practical scenarios. In fact noise, numerical errors and not perfect symmetries in the object itself can all make very difficult for many pairs of points to share exactly the same symmetry plane. One possible solution to this problem could be the use of a threshold in order to state that two plane are equals. Unfortunately this approach introduces two more hurdles: the first is the correct choice for such threshold, which is not different from the choice of the correct bin size, the other is that this way the transitive property would be no more valid, thus breaking the correspondence between cliques in the Consensus Graph and symmetries.

Our proposal is to extend the BSCG presented in 5.2.3 to a weighted graph where each edge is weighted according to the similarity between the planes associated to the vertices connected:

Definition 4. *Given a set of points in space $P = \{p_1, p_2 \dots p_n\}$, their Weighted Symmetry Consensus Graph (WSCG) is the edge weighted undirected graph $G_W = (V_W, E_W, \omega)$ where $V_W = \{(p_a, p_b) \in P | a < b\}$ is the set of pairs of points in P, $E_W = V_W \times V_W$ is the set of edges and $\omega : E_W \to \mathbb{R}^+$ is a weight function that assigns to each edge a positive real value $\omega((p_a, p_b), (p_c, p_d))$ proportional to the similarity between the symmetry plane of p_a, p_b and the symmetry plane of p_c, p_d.*

From a topological point of view a WSCG is by construction a complete graph, so it would not be interesting to search for cliques in it as we made for the BSCG. Instead we are interested in finding large sets of vertices that are all associated to very similar symmetry planes. In other words, since the edges reflect the similarity between planes we are interested in finding sets of vertices that present a high pairwise similarity among them and a low pairwise similarity with respect to vertices external to the set itself. These

two properties (high internal similarity and low external similarity) are commonly used to define the notion of a cluster, which indeed generalizes to the weighted graphs context the notion of clique in binary graphs.

In figure 5.8 we illustrate a simple example of a WSCG. The object shown is a slightly perturbed version of the one shown in figure 5.7. In this case there are no subset of points that exhibit a perfect symmetry, nevertheless the separation plane between a, g is very similar to the one between b, f, which in turn is just a little less similar to the ones between c, d and c, e, and so on (note that only the edges with relevant weight are drawn). The two largest clusters correspond to the two most relevant symmetry planes in the model. It is worth noting than a cluster could assign to the same point in the model two or more different specular points, as in the case of point c which is symmetric to both e and d. This many-to-many property of the assignment is critical when dealing with data coming from sampling, where artifacts and random noise make very unlikely a direct correspondence between each point and its specular instance, even in the presence of perfect symmetry in the sampled object.

5.2.5 Dominant Sets Cluster Extraction

In section 5.2.4 we show how the problem of finding optimal symmetry planes can be casted into the search for large clusters in a set of vertices over which we defined a pairwise similarity measure ω. In order to perform this search we propose to use the Dominant Set framework [123] which is a very effective pairwise clustering technique.

Given an edge weighted undirected graph $G = (V, E, \omega)$, a subset of vertices $S \subseteq V$ and two vertices $i \in S$ and $j \notin S$ the following function measures the coherence between nodes j and i, with respect to the average coherence between node i and its neighbors in S

$$\phi_S(i, j) = \omega(ij) - \frac{1}{|S|}\sum_{k \in S} \omega(ik) \tag{5.4}$$

And overall weighted coherence between i and other nodes in S is defined as:

$$w_S(i) = \begin{cases} 1 & \text{if } |S| = 1 \\ \sum_{j \in S \setminus \{i\}} \phi_{S\setminus\{i\}}(i, j) w_{S\setminus\{i\}}(j) & \text{otherwise} \end{cases} \tag{5.5}$$

Intuitively, $w_S(i)$ will be high if i is highly coherent with vertices in S. Given this measure $S \subseteq V$ is said to be *dominant* if $\sum_{i \in T} w_T(i) > 0$ for any nonempty $T \subseteq S$ and the following conditions hold:

$$w_S(i) > 0, \forall i \in S \tag{5.6}$$

$$w_{S\cup\{i\}}(i) < 0, \forall i \notin S \tag{5.7}$$

The conditions above correspond to the two main properties of a cluster: namely internal homogeneity and external inhomogeneity, which are exactly the properties that we expect to hold for the sets we are searching for.

Another compelling reason to use the Dominant Set framework is that, despite the complexity of their recursive definition, they are very easy to find in practice: in fact a single set (corresponding to a symmetry plane) can be found in polynomial time and a convenient technique exists in order to enumerate all the Dominant Sets in a graph.

Pavan and Pelillo [123] have shown that dominant sets correspond to local maximizer over the standard simplex (i.e. the sets of vectors with components that sum up to 1) of the quadratic function

$$f(\mathbf{x}) = \mathbf{x}^t A \mathbf{x} \tag{5.8}$$

where A is the weighted adjacency matrix of the graph (thus $A_{ij} = \omega(i, j)$).

These maximums can be found by exploiting the convergence properties of the payoff monotonic replicator dynamic

$$x_i(t+1) = x_i(t) \frac{(Ax(t))_i}{x(t)^t A x(t)} \tag{5.9}$$

which is guaranteed to converge to a local maximum when the association graph is undirected and, thus, the matrix A is symmetric[126]. At convergence the value of the function f is a measure of the global coherence of the extracted set, which in turn express the strength of the symmetry found.

Since an object could express more than a single symmetry plane it would be useful to be able to enumerate at least the most relevant Dominant Sets present in the graph. In addition, although replicator dynamics are guaranteed to converge it does not need to stop on the largest Dominant Set, thus by enumerating them we could choose always the most relevant ones. To obtain such enumeration we used the technique proposed in [135].

5.2.6 Experimental Results

In order to show the effectiveness of the proposed approach we performed both qualitative tests on a database of real world objects and quantitative validation over a larger set of synthetic data generated under different conditions of noise and occlusion. In both cases our technique has been compared with the PRST[129] algorithm which we implemented with 64 bins for each dimension of the space of planes parametrization and with ISP refinement. As an optimization meant to keep the size of the Symmetry Consensus Graph reasonable we filtered the vertex set V_W by sampling a fixed number of random vertices.

The qualitative tests have been made against a database of about 40 3D scans of sunglasses models acquired as point clouds with a structured light scanner. Each model counts between 10.000 and 20.000 points which have been sampled with a metric error which is guarantee to be less than 10 micrometers. Our goal is to find for each model the main symmetry plane, that is, in this context, the plane that split each sunglasses pair

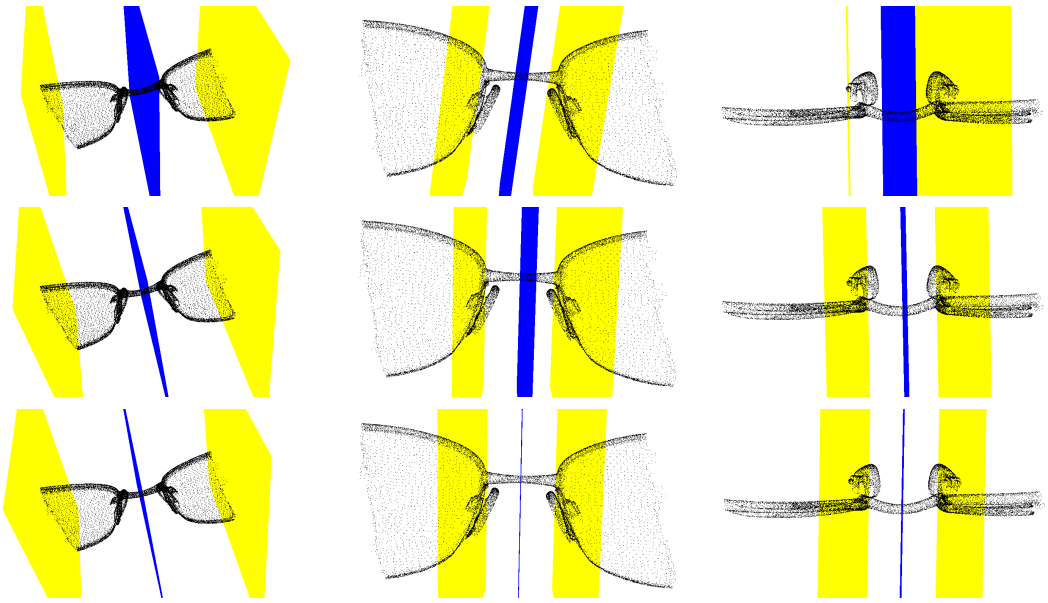
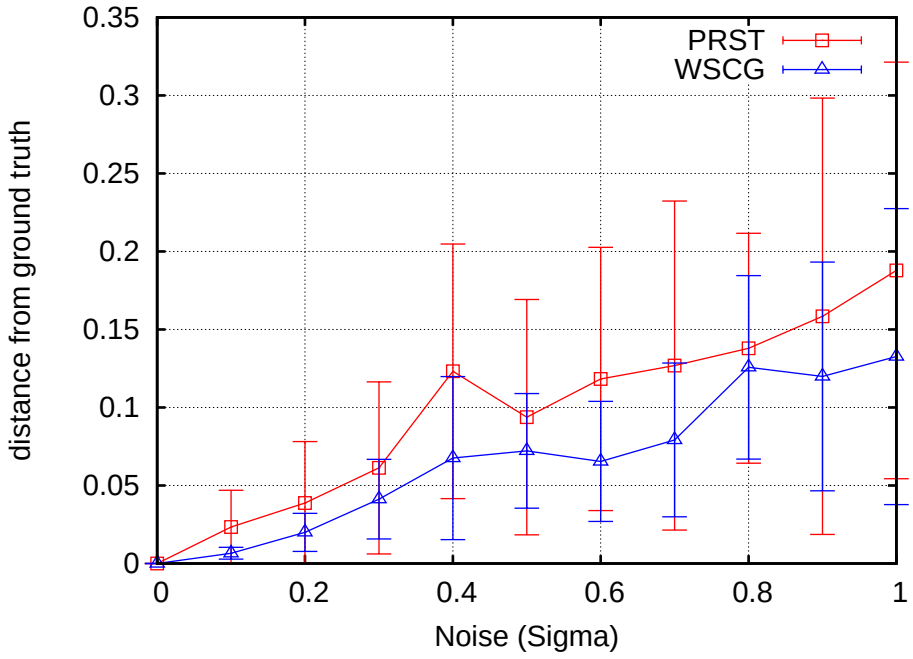


Figure 5.9: Qualitative results on 3D scans. In the first row the symmetry plane obtained with PCA is shown from different angles and details. The second row contains the results of PRST and the third the results of WSCG.

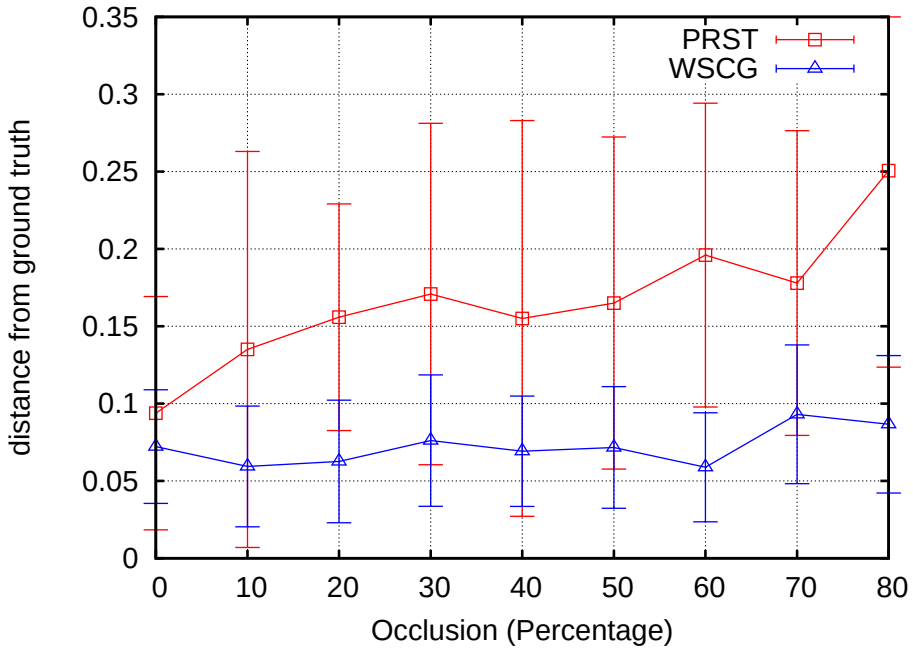
in two specular halves. Despite the high quality of the sampling process it is difficult to obtain an accurate result because of slight differences in the two halves of the model, incomplete sampling and parts of the object that are not fixed and can move during the sample process (for instance nose pads). These hurdles strongly limit the use of traditional global techniques, such as PCA, and the presence of non perfect symmetry can easily lead to some slight imprecision also in more refined methods, such as PRST.

In figure 5.9 we show the performance of these latter methods in comparison to WSCG. The central blue plane is the symmetry plane found, the two lateral yellow planes are added reference planes that are parallel and equidistant to it. The rows in the figure show respectively the plane obtained with PCA, PRST and WSCG. The plane found by PCA is clearly skewed because of the asymmetric sampling of the object during range scanning. The result obtained by the PRST algorithm is far better than PCA, as expected, while WSCG gives another improvement over it.

Qualitative results offer a visual hint about the kind of improvement obtained with the WSCG technique, but as PRST is in itself very accurate this enhancement could appear minimal. In order to give a more quantitative insight we performed a set of experiments with synthetic data which allowed us to compare the symmetry planes obtained with respect to an accurate ground truth plane. We generated a random distribution of 2000 points within a range of 400 millimeters perfectly symmetric with respect to a known plane. To evaluate the performance of WSCG in different conditions we added Gaussian random noise along all three axes with a value of the parameter σ ranging from 0.1 to 1 millimeters. Then we repeated the experiments by setting $\sigma = 0.5$ and simulating oc-



(a)



(b)

Figure 5.10: Accuracy of PRST and WSCG with different level of noise and occlusion

clusion with the deletion of a growing percentage of random points from one side of the symmetry plane. Each experimental condition has been tested with 20 independent runs. In each run the two algorithms are applied and the distance between the obtained princi-

pal symmetry plane and the ground truth is measured. While several symmetry measures have been proposed [185, 98], we are not really interested in knowing how much the found symmetry diverge from a perfect one, as we already know that our data are noisy: rather we are interested in a plane distance measure. Each plane has been parametrized by the vector P perpendicular to it and the distance between two planes is measured as the norm of the difference between their respective associated vectors P_1 and P_2:

$$d = \| P_1 - P_2 \| \tag{5.10}$$

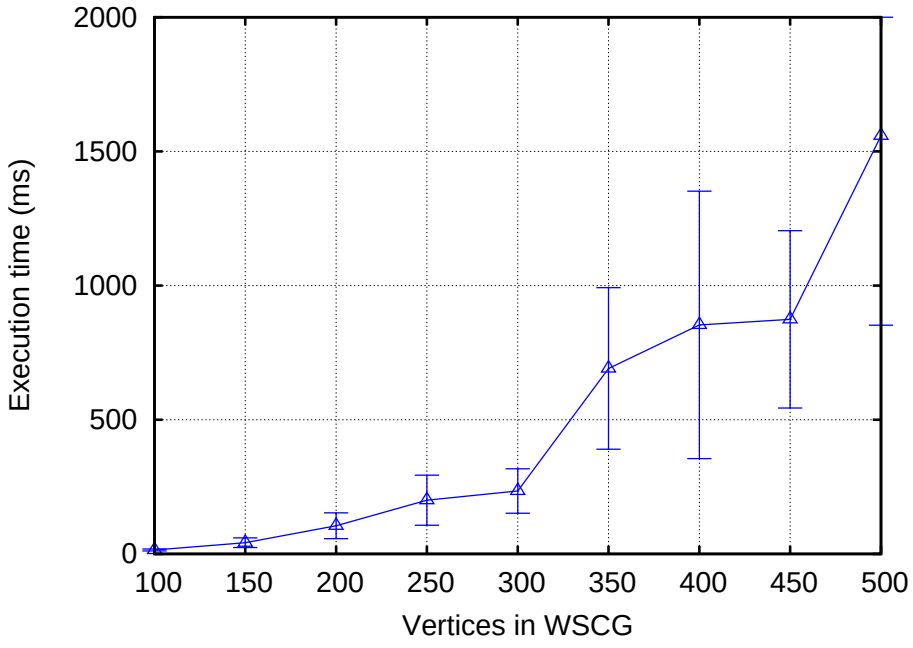
It must be noted that this simple measure mixes up translational and rotational errors, which is reasonable in this context because errors are always very small for both PRST and WSCG. The measure of equation 5.10 is also used to define the function ω that weights the edges of the WSCG in definition 5.2.4. If P_{ab} is the vector perpendicular to the separation plane between points p_a, p_b and P_{cd} the one associated to the plane between p_c, p_d then we measure the similarity between vertices (p_a, p_b) and (p_c, p_d) of the WSCG as:

$$\omega((p_a, p_b), (p_c, p_d)) = e^{-\|P_{ab} - P_{cd}\|} \tag{5.11}$$

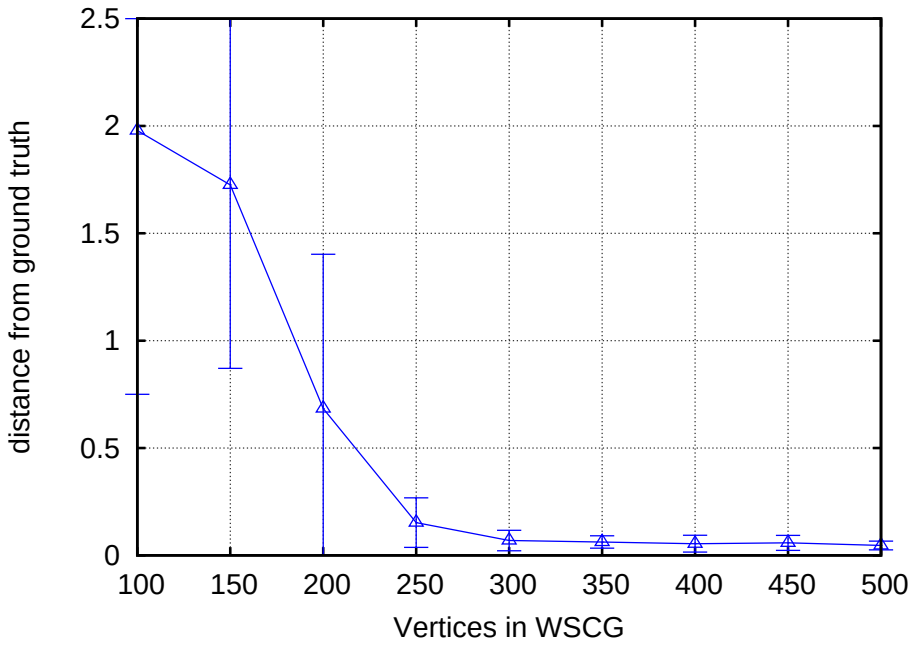
In figure 5.10(a) the performance of WSCG is shown for several levels of Gaussian noise added to the test points and in figure 5.10(b) the robustness with respect to occlusion is shown for different percentages of sample points removal. The comparison with PRST shows that both techniques offer comparable very good performance and that WSCG is more accurate on average. In particular WSCG seems to be more tolerant to occlusion, which is expected as PRST relies on a voting algorithm. In figure 5.11(a) we show the convergence time of the WSCG algorithm with respect to the number of vertices that are kept in the Consensus Graph. Each test has been performed 20 times with Gaussian noise of parameter $\sigma = 0.5$ and an occlusion of 50% and has been executed on a 2Ghz Athlon CPU. These result show that by keeping the number of sampled points low, WSCG can be executed in less than a second. This optimization does not imply a lower quality in the symmetry plane estimation, in fact figure 5.11(b) shows that the error becomes very small when keeping as few as 300 vertices.

5.2.7 Concluding Remarks

The Weighted Symmetry Consensus Graph is a new approach to planar symmetry detection that exploits pairwise clustering to find a solution that is maximally consistent from a global perspective. The main features of WSCG are that it does not require any quantization of the solutions space and it is adaptive with respect to outliers. Experimental validation shows that this approach gives better results than another state-of-the-art algorithm. Finally WSCG is quite general as the compatibility measure needs not to be related to planar symmetry. In fact future work will aim to extend the approach to radial symmetry and other features that can be estimated by exploiting a global consensus.



(a)



(b)

Figure 5.11: Running time and accuracy of WSCG with respect to the number of points

5.3 Unambiguous Compound Phase Coding for Surface Reconstruction

Phase shift methods have proven to be very robust and accurate for photometric 3D reconstruction. One problem of these approaches is the existence of ambiguities arising from the periodicity of the fringe patterns. While several techniques for disambiguation exist, all of them require the projection of a significant number of additional patterns. For instance, a global Gray coding sequence or several supplemental sinusoidal patterns of different periods are commonly used to complement the basic phase shift technique. We propose a new pattern strategy to reduce the total number of patterns projected by encoding multiple phases into a single sequence. This is obtained by mixing multiple equal-amplitude sinusoidal signals, which can be efficiently computed using inverse Fourier transformation. The initial phase for each fringe is then recovered independently through Fourier analysis and the unique projected coordinate is computed from the phase vectors using the disambiguation approach based on multiple periods fringes proposed by Lilienblum and Michaelis[100]. With respect to competing approaches, our method is simpler and requires fewer structured light patterns, thus reducing the measurement time, while retaining high level of accuracy.

5.3.1 Structured Light Coding Strategies

The main challenge for any triangulation-based surface reconstruction technique is the assignment of reliable correspondences between features observed by two or more different points of view. Given the central role of this problem, many and diverse strategies have been proposed in literature over the past few decades [139]. When a sparse surface reconstruction is adequate, correspondences can be searched and tracked among repeatable features readily present in the scene, such as corners or edges. Unfortunately, in general it is not possible to guarantee that the same features are extracted from each image, or that the feature density is sufficient. Hence, complementary techniques, usually based on photometric correlation, are used to obtain an approximate reconstruction of the scene depth map. Structured light systems overcome these limitations as they do not rely on natural features, but instead use projected patterns of light in order to find correspondences that are usually as dense as the pixels of each image [87]. Among these, time-multiplexing strategies such as n-ary and Gray codes, as well as hybrid approaches, are by far the most utilized [88].

Simple binary coding assigns to every pixel a codeword retrieved from the digitized sequence over time of projected black and white stripes; binary coding methods require $log_2(t)$ pattern images to generate t code strings. Robustness of binary codes is improved by using Gray codes, where adjacent codes differ only in one bit. Both the techniques generate unique codes along each scanline, but at the same time are limited by their low resolution due to the inherently discrete nature of the coding. Also, the large number of projected patterns does not result in an increased accuracy. Generally, this class of

measurements proves to be ineffective with objects having different reflective properties (such as slick metal parts or low reflective regions), thus they must rely on the assumption of uniform albedo [88].

Phase shifting methods, on the other hand, yield higher resolutions since they are based on the projection of periodic patterns with a given spatial period. Each projected pattern is obtained by spatially shifting the preceding one of a fraction of the period, and then captured by one or more cameras. The images are then elaborated and the phase information at each pixel determined by means of M-step relationships [153]. Since the phase is distributed continuously within its period, phase shifting techniques provide subpixel accuracy and achieve high measurement spatial resolution. Furthermore, the intensity change at each pixel for subsequent patterns is relative to the underlying color and reflectance, which makes phase shift locally insensitive to texture variance to a certain degree. A major drawback is that, in its basic formulation, phase shifted structured light renders only relative phase values and thus it is ambiguous. However, when both an extended measuring range and a high resolution are required, a combined approach proves to be very powerful. The integration of Gray code and phase shift brings together the advantages of both, providing disambiguation and high resolution, but the number of patterns to be projected increases considerably, and each strategy introduces a source of error [100].

Other high resolution shape measurement systems include optical profilometers. Non-contact phase profilometry techniques relate each surface point to three coordinates in a frame having the z axis orthogonal to a reference plane, which then represents the reference for the measured height [147, 178]. In classical phase measurement profilometry, gratings or sinusoidal patterns are projected and shifted first onto the plane and then over the object to be measured. Phase information from the deformed fringe pattern is then extracted by means of various techniques. Other, more effective profilometry techniques include the well-known Fourier Transform method [154] and other interesting derivatives [183, 73]. Fourier-based profilometry can require as few as one or two frames for depth estimation, which makes real-time reconstruction possible. Nevertheless, in profilometric methods phase variation caused by height modulation must be limited: ambiguity of the phase limits the measurement range, allowing for gauging of smooth-shaped objects only. Moreover, noise from camera, distortion of lens, difficulties of calibration, aliasing and imperfectness of the projecting unit influence the precision of most of these techniques [40, 149].

Recently, some number theory based methods have been proposed for disambiguation [190, 100]. In [100] the authors relate absolute, unambiguous phase values to projector coordinates $\xi \in \mathbb{R}$, and define $\xi(u, v)$ to be the projector coordinate at pixel (u, v). Then several phase shift sequences, each with a different local period λ_i, are projected onto the object to be measured. A phase image is obtained for each sequence through the computation of periodic phase values $\phi_i(\xi) \in [0, 1)$ at every pixel. In addition, the fringes of a pattern are assigned sequential natural numbers $\eta_i(\xi) \in \mathbb{N}$, which represent a simple counting of the fringes from left to right. A projector coordinate can then be directly

obtained, for all $i = 1, 2, ..., n$, from a fringe number and a phase value:

$$\xi = (\eta_i(\xi) + \phi_i(\xi))\lambda_i \,. \tag{5.12}$$

Since the only available values during measurement are λ and ϕ, it is clear that the system of equations becomes ambiguous as the same value of ξ can be obtained for different values of η_i. This happens when two different projector coordinates yield the same phase values for all i. Therefore, the authors follow a number-theoretic approach and identify a general condition for generating unambiguous pattern sequences, by defining a maximum projector coordinate ξ_{max} up to which ambiguity can be excluded. Such a coordinate is defined as the least common multiple of relatively prime periods λ_i, and clearly for practical advantage it must entirely cover the projector range. An efficient method is given to calculate the fringe numbers from the ambiguous phase values at each pixel, given the local period lengths. This method takes advantage of a simple relationship between phase values and fringe numbers. Given any pair of pattern sequences, the following equivalence holds for each image pixel:

$$\lambda_i\phi_i(u, v) - \lambda_j\phi_j(u, v) = \lambda_j\eta_j(u, v) - \lambda_i\eta_i(u, v) \,. \tag{5.13}$$

This makes it possible to construct a theoretical phase difference vector beforehand, and then use it to retrieve the fringe numbers when real phase measurements become available. In addition to providing an efficient way to obtain the fringe numbers, this method allows to assign each point a reliability value related to the deviation between measured and expected values. The use of theoretical phase difference vectors makes for a powerful test, which allows to identify erroneous or weak measurements (such as mixed phase values) caused, for instance, by sharp edges, involuntary object movements and light reflections. Once the unknown fringe numbers are calculated, projector coordinates can be easily retrieved for each pattern sequence with equation 5.12. The independent measurements can then be averaged to obtain a unique and absolute phase value at every pixel in an efficient way, leading to an increase in accuracy of the measurements.

The big advantages of the multi-period method is its relative simplicity and high efficiency. The phase-coded images can be directly employed in general stereo reconstruction systems, ensuring high quality and density of the code. Specifically, the lack of surface points is mainly due to occlusions and camera disparity, and measurement errors are very low thanks to the averaging and validation procedures implicit to the approach, that exclude a large percentage of errors and outliers before the actual surface reconstruction takes place. The main drawback lies in the fact that, typically, three or more pattern sequences are needed to entirely cover the projector range (typical values are 800 or 1024 projector pixels). This requires the projection of as many as three times more patterns than required with classical phase shifting.

In the next section we introduce a novel coding strategy that retains the big advantages offered by the multi-period method, but it requires a significantly lower number of structured light patterns while achieving comparable levels of accuracy.

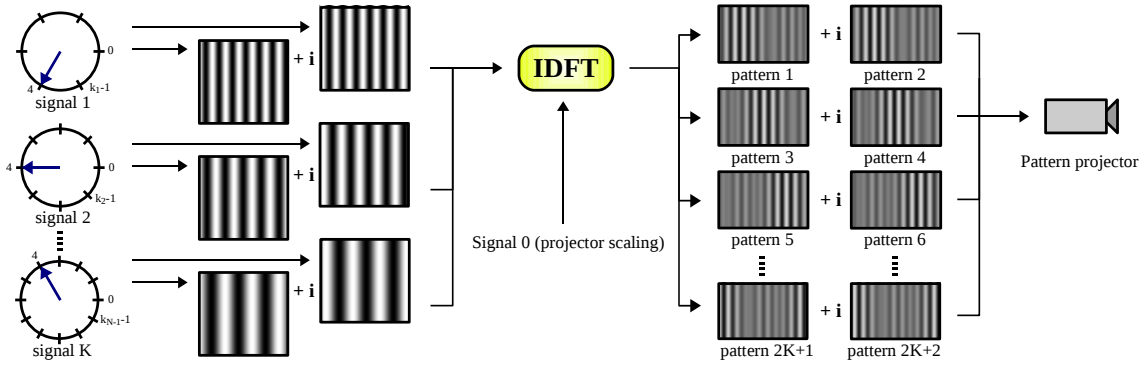


Figure 5.12: The composition of k fringe patterns, plus one unknown shift signal will produce a total of $2(k+1)$ image patterns that will be projected onto the surface to be reconstructed. The shift pattern (projector scaling) accounts for the unknown value of the albedo of the surface.

5.3.2 Compound Phase Coding

The main idea behind the Compound Phase Coding strategy is to project several fringe patterns in a single spatio-temporal pattern. This is obtained by encoding the phases of the fringe vector as phases of a Fourier term at different frequencies. Each fringe is characterized by a different period and all of them are relatively prime. This way, once recovered the single initial phase shift for each fringe, we are able to build an unambiguous code with a numerical technique similar to the one suggested in [100].

Let $\lambda_1, \ldots, \lambda_k$ be k periods and let ξ be the projector coordinate at some pixel. If the periods are coprime and the projector coordinates do not exceed $\prod_{j=1}^{k} \lambda_j$, then we have a unique phase code for ξ [100]. Namely, this code is given by the vector $\boldsymbol{\phi} = \begin{pmatrix}\phi_1 & \ldots & \phi_k\end{pmatrix} \in [0,1)^k$, where $\phi_j = (\xi \mod \lambda_j)/\lambda_j$.

Our aim is to map $\boldsymbol{\phi}$ into a signal sequence of gray-scale values that can then be measured by the cameras to obtain the unique code of each fringe. We take the hint from phase shift methods that phase encodings are more robust than amplitude codings, but extend multi-period coding by decoupling the phase used to encode the message from the frequency of the sinusoidal signal used to transport it. This is done by projecting the sum of equal-amplitude sinusoidal signals at frequencies $\frac{1}{k+1}, \frac{2}{k+1}, \ldots, \frac{k}{k+1}$, where k is the number of periods necessary to encode the coordinate x_i, and by encoding the phase parameters as the phases of the corresponding sinusoidal signal.

Given a phase code $\boldsymbol{\phi} \in [0,1)^k$, we create a $(k+1)$-dimensional complex vector $\mathbf{x} \in \mathbb{C}^{k+1}$, where

$$x_j = \begin{cases} 0\,, & \text{if } j = 0\,, \\ e^{-2\pi i \phi_j}\,, & \text{if } 1 \le j \le k\,. \end{cases}$$

Here, $i = \sqrt{-1}$. Note that given x_j for any $1 \le j \le k$, we can compute the phase ϕ_j as

$$\phi_j = \mathrm{frac}(1 + \frac{1}{2\pi} \arg(\Im(x_j), \Re(x_j)))\,, \tag{5.14}$$

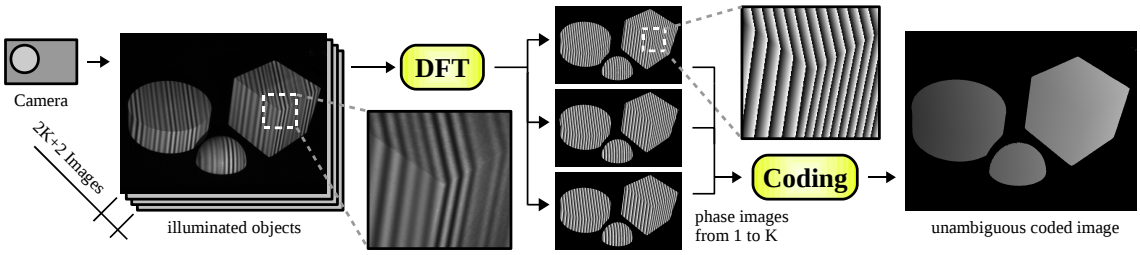


Figure 5.13: A total of 2K+2 images of illuminated objects are captured and single phase values are calculated for each composed fringe signal. Those values are subsequently used to get an unambiguous coding. Note that the intensity profile of each projected pattern is not sinusoidal.

for any value of s, where $\mathrm{frac}(\cdot)$ is the fractional part of the argument, and $\Im(z)$, $\Re(z)$ are the imaginary and real parts of $z \in \mathbb{C}$, respectively.

Each complex number x_j represents the amplitude and phase of a sinusoidal component with frequency $\frac{j}{k+1}$ cycles per sample. Hence we can reconstruct the intensity sequence of that coordinate by computing the Inverse Discrete Fourier Transform of $\mathbf{x}$, obtaining the vector $\mathbf{y} \in \mathbb{C}^{k+1}$, where

$$y_n = \frac{1}{k+1} \sum_{j=0}^{k} x_j e^{2\pi i \frac{j}{k+1} n}, \quad n = 0, \ldots, k,$$

.

We can then project separately the real and imaginary part of this vector as two time sequences obtaining a single set of $2(k+1)$ patterns to be projected to uniquely encode the x_i projector coordinate (see Figure 5.12).

Hence, given $\mathbf{y} \in \mathbb{C}^{k+1}$, we transform it into a real vector $\mathbf{z} \in \mathbb{R}^{2(k+1)}$, where $z_{2j-1} = \Re(y_j)$ and $z_{2j} = \Im(y_j)$ for $1 \leq j \leq k+1$. Afterwards we scale and shift each component of $\mathbf{z}$ in order to bound them within $[0, 255]$, obtaining the sequence of gray-scale values to project in correspondence to ξ. Note that by applying a shift we affect the information at frequency 0, while by scaling, we modify the amplitudes of all frequencies, without influencing the phase values.

The acquisition process introduces an additional linear deformation on $\mathbf{z}$, which depends on the physical properties of the object being scanned. Again this does not affect the phases.

Let $\overline{\mathbf{z}} \in \mathbb{R}^{2(k+1)}$ be the acquired gray-scale values and let $\overline{\mathbf{y}} \in \mathbb{C}^{k+1}$ be its representation into a complex vector. Then, the net effect of the projector scaling and the change in the reflectivity properties of the surface is a translation and scale of the observations, i.e. the relation between the intended signal $\mathbf{y}$ and the observed signal $\overline{\mathbf{y}}$ is

$$\overline{\mathbf{y}} = \overline{\delta} + \overline{s}\mathbf{y}$$

for some real values $\overline{\delta}$ and $\overline{s}$. The phase code is finally recovered from $\overline{\mathbf{y}}$ by computing

the Discrete Fourier Transform, namely

$$\overline{x}_j = \sum_{n=0}^{k+1} \overline{y}_n e^{-2\pi i \frac{n}{k+1} j} ,$$

and by applying (5.14) to the result $\overline{\mathbf{x}} \in \mathbb{C}^{k+1}$, obtaining $\boldsymbol{\phi}$ (see Figure 5.13).

This process allows to recover the phase code for each projector coordinate by taking only $2(k+1)$ measurements, where k is the number of signal periods. Nevertheless, one can also force a larger number of samples in order to increase accuracy, by appending null components to $\mathbf{x}$. More precisely, by appending M null components, we need $2(M+k+1)$ measurements in order to recall the phase code $\boldsymbol{\phi}$.

It should be noted that a drawback of this approach is that encoding multiple signals in a single pattern reduces the effective projector intensity range available to encode each phase, increasing the effects of the discretization error. However, experiments show that the error introduced by this is limited.

5.3.3 Experimental results

In order to validate the proposed technique we need to compare its accuracy to the results obtained with a state of the art phase shift technique. To this extent we choose to compare our measurements with those given by the Multi-Period Phase Shift proposed in [100]. The reasons for this choice are two-fold: Multi-Period Phase Shift is very accurate, as it uses information for each fringe projected in order to reduce the average error; in addition, once the phase vector from the composite signal is obtained, measurement quality is directly comparable as both techniques share the same numerical disambiguation step. In the following sections we will show both quantitative results, by evaluating the measurement error over a planar target, and qualitative results, by showing relative average distances between the measurements obtained by the two techniques with generic objects and by comparing the corresponding estimated quality.

Experimental setup

All the following experiments have been run on a test rig for structured light techniques that has been internally developed in our lab (figure 5.14). The rig is made up of a motorized plate for object positioning, four cameras and an illumination source mounted on a motorized liftable platform. Specifically the cameras are equipped with a 1/2 inch CMOS sensor which offers a full 1280x1024 resolution. The cameras are monochrome, thus no Bayer filters are placed over the sensor. While four cameras are available, in this experiment set we use only one pair of cameras to reconstruct the surfaces. Thus the system falls into the category of two calibrated cameras and one uncalibrated light source. The illumination source is a 800x600 color DLP projector which we use to project the monochromatic patterns. The system is controlled by a standard PC housed into the base of the rig. This PC is a 2.8 GHz AMD quad core system with 2 Gigabytes of ram.

Intrinsic and extrinsic parameters of the cameras have been obtained through a standard calibration procedure using a planar checkerboard and OpenCV[31] calibration software. It must be noted that our system is not a full fledged production scanner, thus it does not guarantee extreme accuracy or resolution: in fact we estimate its precision in about $30\mu m$. This level of accuracy is adequate with respect to our experiments, as we are not interested in showing the absolute precision of our setup, but rather the relative performance of the coding schemes, showing that the proposed approach can be used as an effective replacement for slower approaches without suffering from a significant loss in accuracy.

Planar target measurements

In this first set of experiments we measured the surface of a 200 by 200 mm squared piece of float glass which we previously sprayed with a very thin layer of acrylic paint. We made several sets of measurements with both the Compound Phase Coding technique and the Multi-Period Phase Shift technique. Since the exact pose of the test object is unknown and the surface cannot be perfectly flat, we approximated the ground truth with the best

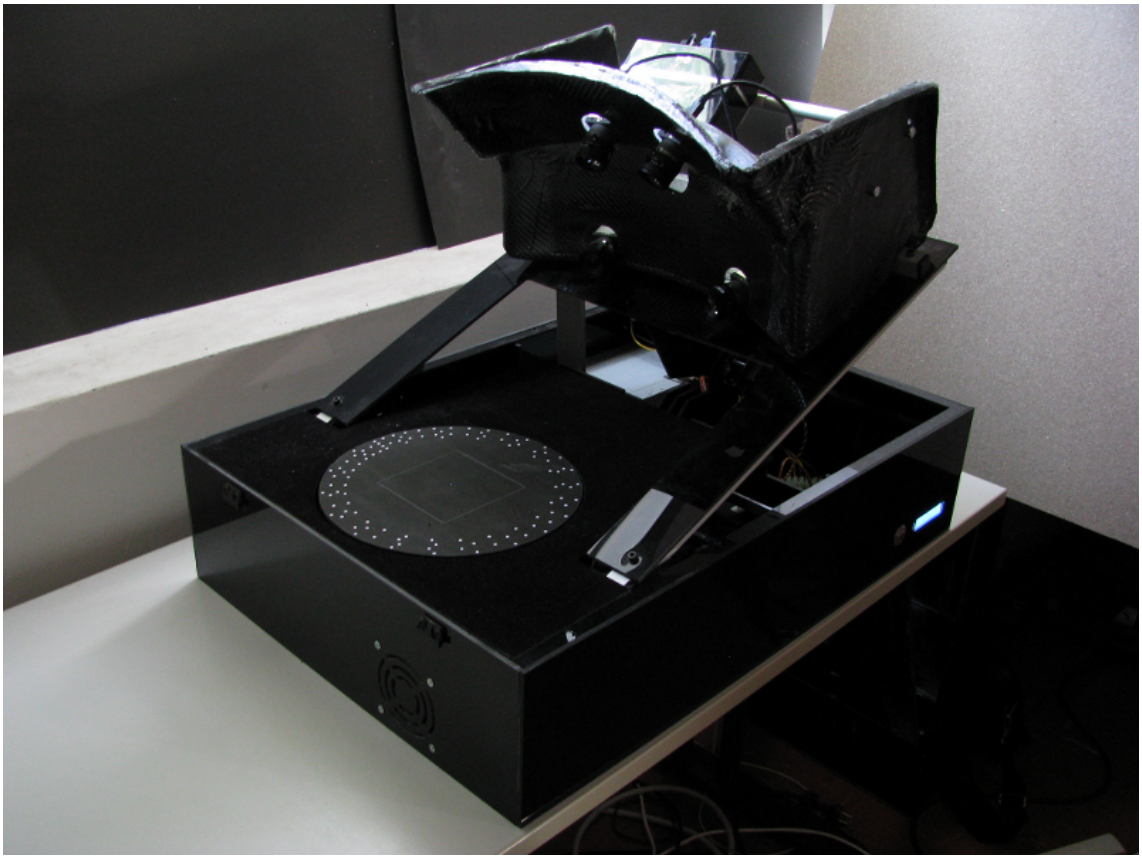
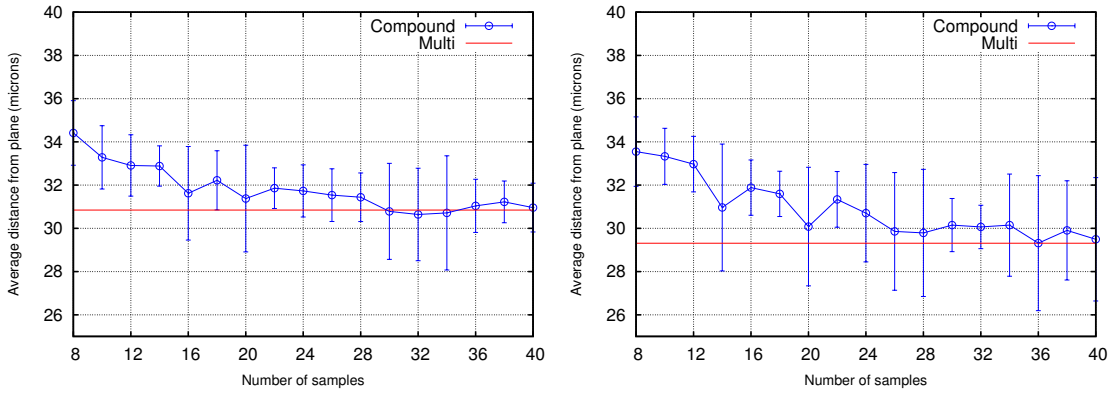


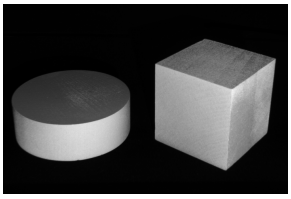
Figure 5.14: The general purpose structured light scanner used.



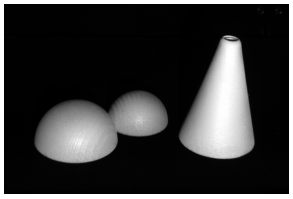
(a) (b)

Figure 5.15: Accuracy comparison between the Compound Phase Coding method and the Multi-Period Phase Shift technique. In figure 5.15(a) we used periods of 7, 11 and 13 pixels (30 patterns for Multi-Period Phase Shift), and in figure 5.15(b) we used periods of length 9, 11 and 13 (34 patterns for Multi-Period Phase Shift). Note that the Multi-Period technique appears as a flat continuous red line and its standard deviation as dashed red lines. Vertical bars are standard deviations in the measurement of the error for the compound technique.

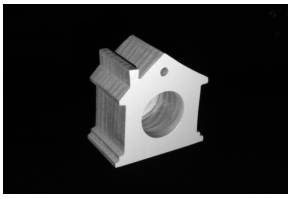
fitting plane (in the least squares sense) with respect to the measured points of the object. This way we estimate the expected measurement error of each technique as the average of the absolute value of the distance from the fitting plane of each measured point. We had a wide range of choices regarding the number of different signals to project and their periods. We chose to execute the test with two configurations: 3 signals of periods respectively 7, 11 and 13 pixels and other 3 signals of periods 9, 11 and 13 pixels. Since our projector has an horizontal resolution of 800 pixels both configurations allow to obtain a globally unambiguous coding of the object. Given those signal configurations, we projected respectively 30 and 34 patterns for testing the Multi-Period Phase Shift technique. Since we always used 3 signals, the Compound Phase Shift technique strictly requires only 8 patterns to be projected: nevertheless we repeated the measurement with a growing number of additional patterns in order to study the effect of the supplementary information on the final accuracy. Finally, each experimental measure was repeated for a total of 10 times. In figure 5.15 we compare the performance of the proposed approach against the baseline multi-period approach. Accuracy of the multi-phase approach is slightly better in 5.15(b), probably due to the higher number of patterns projected (34 instead of 30). The blue line shows the trend of the average error of the proposed Compound Phase Coding technique as the number of projected pattern is increased. It should be noted that even with the minimal number of projected patterns the accuracy is quite good: in fact, on average our distances from the fitting plane are only about three micrometers higher than those obtained using the Multi-Period technique, which requires almost four times as many patterns. Moreover, by projecting additional patterns, the qual-



Cube and Disc



Cone and Spheres



Toy House

Technique	Multi	Cmp8	Cmp16	Multi	Cmp8	Cmp16	Multi	Cmp8	Cmp16
Points	109347	108306	109034	46749	46263	46505	37457	36880	37133
Avg dev.	0.025	0.089	0.063	0.030	0.085	0.067	0.036	0.088	0.068
Avg dist.	-	32.00	25.09	-	21.15	16.19	-	23.26	20.87

Figure 5.16: Comparison between reconstructions obtained with the Multi-Period Phase Shift and the Compound Phase Coding on several test objects. Compound Phase Coding has been tested using both 8 and 16 samples. Multi-Period Phase Shift has been tested with 34 samples in order to obtain the best quality. Distances are in microns and objects are 5 to 10 cm wide.

ity of the measurements can be further enhanced: for instance, by doubling the number of projected patterns the distance from the Multi-Period approach is approximately halved. Finally, as we expected, using the same number of patterns projected by the Multi-Period technique, the two approaches yield equivalent performances. Of course, the whole point of the proposed technique is to reduce the number of projected patterns. In this sense, our approach allows to control the time/precision trade-off, obtaining good reconstructions even with just 8 patterns.

Generic objects measurements

In this set of experiments we measured the surface of several generic scenes built with simple wooden objects in a volume of about 200 mm of diameter in each direction. The set of objects was selected to maximize the range of surface orientations and conditions. Again we compared our results with those obtained by the Multi-Period approach: in this case we take the Multi-Period results as a direct comparison, as no ground-truth or knowledge of the surface of the objects (which are hand-made) is available. Compound Phase Coding was tested both with 8 and 16 projected patterns. Since we projected signals with periods of respectively 9, 11 and 13 pixels Multi-Period Phase Shift was tested with 34 samples. For each experiment we evaluated three quantities, namely:

- The number of points acquired. That is the number of surface points that pass both the consistency and the quality check. The first verifies that the phase vector is consistent with the code assigned to the point, the latter ensures that the phase difference vector contains only integer values (as expected with respect to constraint 5.13). Specifically, for these experiments we rejected points associated to phase

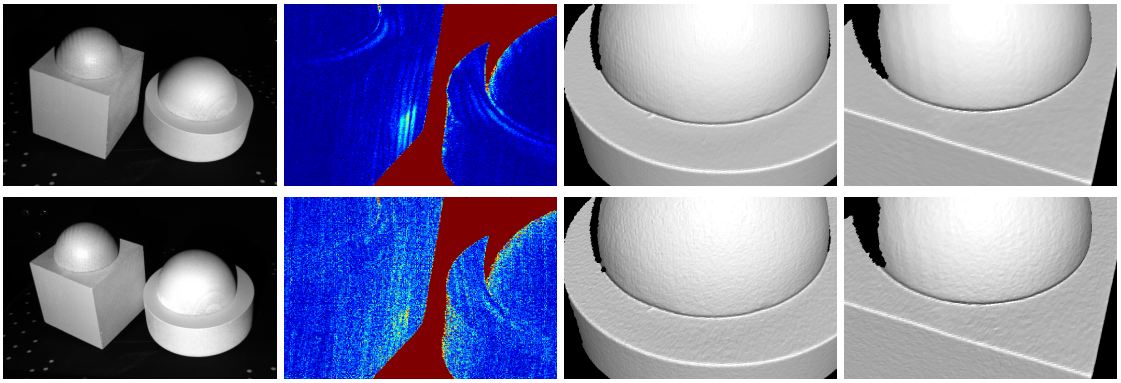


Figure 5.17: A surface reconstruction example complete with deviation assessment. In the first column, the object as viewed by the left and right camera. In the second column, the deviance maps of Multi-Period Phase Shift (top) and Compound Phase Coding (bottom). In the last columns, two respective close-ups of the reconstructed surfaces.

difference vectors where at least one entry deviates from the nearest integer by more than 0.2;

- The average deviation. That is the average distance from each entry in the phase difference vector and the nearest integer. Since each element of the phase difference vector should be integer a lower value suggests a higher quality in the measurement obtained. This parameter has been calculated over all the points obtained, even those filtered by the consistency and quality checks;

- The average distance. That is the average of the absolute distances between points obtained by the Multi-Period and Compound techniques. Note that this measure is possible because we implemented both algorithms in a way that allowed us to produce depth maps that are exactly overlapped along the x and y axis and differ only for the depth values assigned along the z axis.

In Figure 5.16 we report the values of those three quantities in different scenes. In general, the number of points acquired by the Compound Phase Coding with only 8 patterns is slightly smaller than the number of valid points given by the Multi-Period technique. This distance, while small, is almost completely eliminated when using 16 patterns. The average deviation resulting by applying our technique is always slightly higher, nevertheless it should be remarked that those values are all quite small: in fact, for any practical purpose a deviation smaller than 0.1 can always be associated with a valid measured point. Finally, the average distances between the two techniques also assess the suitability of our technique as a fast replacement for traditional Multi-Period: in fact, the results obtained are generally compatible with the measurement error estimated in the experiments done in section 5.3.3, a result that should be expected since both techniques are subject to measurement errors. It should also be noted that the higher average distance obtained with the scene “Cube and Disc” is probably due to the presence of sharp edges and of two large

surfaces at a grazing angle with both the cameras and the light source. Conversely, the higher smoothness offered by the scene "Cone and Spheres" allows for a more precise reconstruction.

In Figure 5.17 we show a more qualitative example of the difference in accuracy of reconstruction between our method and a complete Multi-Period phase shift. In this experiment we used the less accurate instance of Compound Phase Coding, thus only 8 patterns were projected. In the first column we reproduced the scene viewed from the left and right camera. In the second column we calculated a map of the maximum deviation from integer values of the phase difference vectors of the Multi-period method (top) and the Compound method (bottom). In these maps we used the standard Matlab color scale, thus a full blue pixel indicates a low deviation and a bright red pixel a maximum deviation (in this case 0.5). Uncoded pixels are represented in dark red. Both techniques exhibit a fairly low deviation, with high values mainly on the borders, which is due to the camera integration of the projected signal and the background on boundary pixels. The other high deviation area on the side of the cube is probably due to inter-reflection between the two objects.

The last two columns show a close-up of the surfaces reconstructed with both methods. In general the surface reconstructed with the Compound technique is a bit more noisy, but the overall level of detail is maintained, even with details at a relatively high frequency being clearly visible on both reconstructions: for instance, in the close-up of the disc object (third column) a very small bulge is visible in both the reconstructions in proximity of the lower rim of the hemisphere.

Finally, in Figure 5.18, we compare the reconstructed phases obtained after the application of the Discrete Fourier Transform to the captured structured light images before and after the coding step. As in the previous experiments we choose periods of 9, 10 and 13 pixels and we projected 34 patterns for the Multi-Period method and 8 pattern for our approach. In the first column we show respectively the complete coding for the Multi-Period (top) and for the Compound (middle) methods. In order to highlight the differences between the two codings, we show in the third line of this column the difference image between them: as expected they are almost identical and the only slight differences are found in boundary regions, where the measurements are less precise.

In the other two columns we show the phase related to two different signals calculated by the two techniques (we show only two signals for space reasons). For the purpose of a simple comparison between them, we plotted a magnified graph of the small red lines in the phase images. In those graphs the red line represents the phase extracted by the Multi-Period method and the blue line the phase extracted by our technique. It can be seen that the values obtained are very similar, and that they diverge by a very small amount only in proximity of critical areas such as the interface between the disc and the hemisphere.

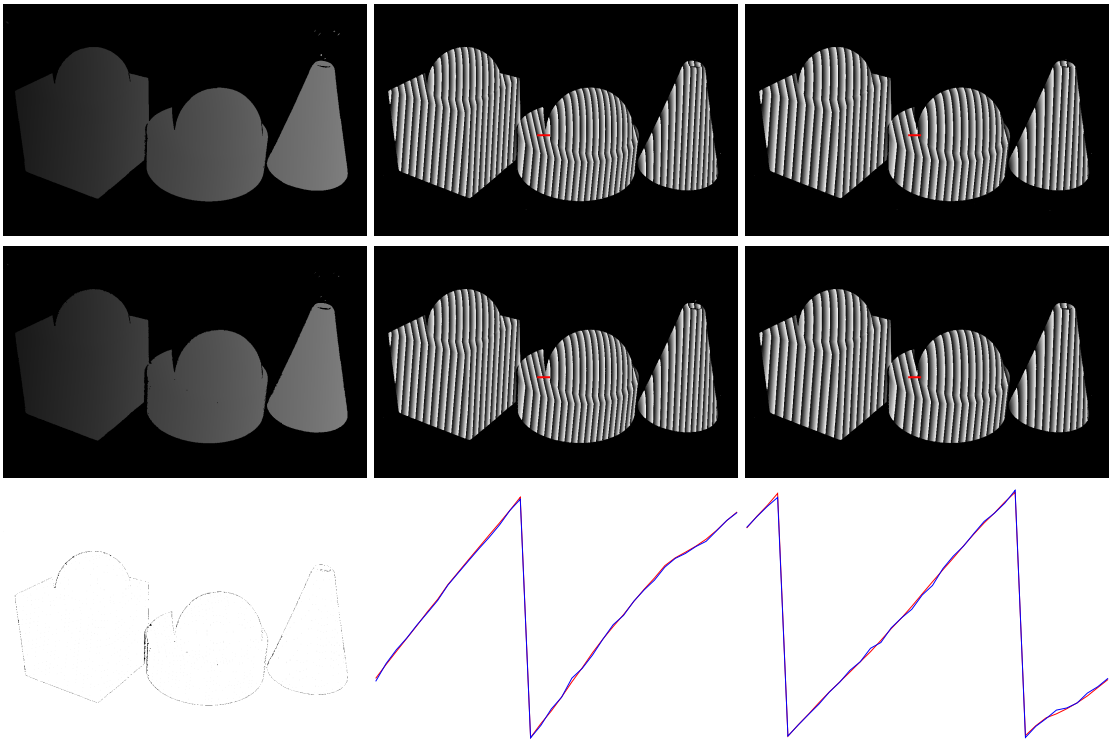


Figure 5.18: A comparison of the coding and of the phase values extracted by our technique and by Multi-Period phase shift. In the first column we show the final codings and their difference. In the other columns we display the respective phase images for two reconstructed signals. The two magnified plots in the last row correspond to the small red segment. Note that the obtained phases are very similar even in the proximity of critical areas.

5.3.4 Concluding Remarks

In this section we have proposed a novel compound phase coding technique that is able to perform a complete and accurate surface reconstruction requiring the projection of as few as 8 patterns. This is obtained by encoding multiple phase information in a single pattern sequence as phases of sinusoidal signals at integer frequencies, thus encoding and decoding the compound signal using standard Fourier analysis. The comparison with another well known technique assesses the ability of the approach to obtain accurate reconstruction even with very few patterns. In addition, the time/quality trade-off can be easily controlled by adding more patterns. In fact, we show that measurement error decreases consistently by adding more information, to the point that our method reaches the performance of other state-of-the-art approaches when feeded with a comparable quantity of data. Finally, while a comparison with commercially available scanners is out of the scope of this research, we can observe that we obtain an average deviation from the surface of about 30-40 microns, which is on par with the precision offered by high end commercial scanners.

6
Conclusions

In this thesis, we introduce a novel game-theoretic based framework for matching and inlier selection. We tested its effectiveness over a wide selection of applicative domains. For each scenario we propose a special crafted adaptation that allows for a better handling of domain-specific features. To this extent, the general framework is extended in each addressed context with appropriate theoretic and experimental validation.

Contributions and novelty of the work

In Chapter 3 we introduce the basic framework and test its performance in solving matching problems by comparing its results with those obtained by standard RANSAC-based inlier selection techniques. Specifically we define the concept of matching game and the modeling of matching candidates as pure strategies. We show how to enforce hard constraints and thus how to adapt the framework to handle both one-to-one and one-to-many matching. Finally we suggest the use of evolutionary dynamics to evolve an initial population of strategies toward an evolutionary stable state. In Chapter 4 we offer two different approaches for matching data sets where high order properties need to be enforced by each group of three or more matched items. Specifically we introduce the use of high order attributed graphs to solve this type of problems. To this extent, a maximal weight subgraph isomorphism is searched in a special crafted association graph by using first a reactive local search scheme (in section 4.1) and evolutionary game dynamics (in section 4.2). Both approaches exceed the state of the art in their respective field of application. Further we propose an complementary view of the problem, by using an association hypergraph between single matches rather than build a standard association graph with higher order tuples as nodes. This latter approach is slightly more scalable with respect to the order of the enforced property and offers a comparable level of performance. In Chapter 5 we propose two applications to 3D reconstruction. In the first section we show how to use our game-theoretic approach for coarse and fine surface registration without requiring any form of initial motion estimation. Here we define a very detailed matching technique which is based on sound theoretical properties. In addition the experimental validation shows very good results: in particular we always obtain the correct registration without falling in local minima and the quality of the alignment itself is comparable

with that obtained by state-of-the art iterative registration techniques. In section 5.2 we show how to use the proposed approach to find symmetry planes in point clouds. Here we introduce an association graph-based problem formulation which in turn uses evolutionary techniques in order to search for the best solution. Experimental results shows a higher precision with respect to other commonly used techniques. Finally, in section 5.3 we propose a novel coding technique for structured light 3D reconstruction. While not game-theoretic related, we obtained this interesting result as a by-product of our research on 3D structured light scanning technologies. We believed relevant to include it in this thesis as it is integral with the whole research path followed.

Strengths and weaknesses

Our approach exhibits three main advantages over most inlier selection techniques. The first advantage is the ability to exploit a global coherence model on the basis of local compatibilities expressed between small sets of two or few more matching candidates. This is the main feature that distinguishes our approach from the current literature. In fact, this allows to leverage on the transitive closure of the property over all the sets and to guarantee its enforcement among all the members of the final configuration. Furthermore, the approach offers a high level of flexibility: by choosing different payoff function it is possible to attack very different problems, in addition it is easy to switch between one-to-one and many-to-many matching approaches. The second advantage is the ability to deal effectively with feature descriptors that show poor distinctiveness. This is quite evident in the image and surface matching applications, where correct matches are detected with high precision, even in the presence of several feasible candidates for each match. The third advantage is that our technique does not require to fix any threshold. We think that the thresholding is the main hurdle with most of other approaches: in fact it is very difficult to select an appropriate value in advance and without knowledge of the correct solution. By contrast, our technique reacts appropriately to different conditions of the data by auto adapting in virtue of a sort of automatic thresholding driven by the global average payoff of the currently active population.

During our work we found two main drawbacks in the proposed technique. The first is of course the fact that it is not able to model all of the matching problems. In fact we are only able to attack problems where a global property is verifiable over a subset of two or more matches. While this is not the whole set of scenarios, this is the case for most parameter estimation problems, as well as structural constrained problems. The second drawback is related to the computational and spatial complexity of the framework. Since each match candidate is modeled as a strategy, the size of the payoff matrix between strategies grows with the square of the match candidates. In typical problems match candidates can easily be in the order of thousands, leading to data sizes in the order of several millions of payoffs. This can be a drawback both computationally (since most dynamics need to perform several iteration of matrix-vector multiplications) and spatially (since payoff matrices can grow up to several hundreds of megabytes). Fortunately this

drawback is strongly moderated by the use of infection/immunization dynamics, which allows for both faster convergence and for off-line storage of the data, since only one column of the payoff matrix is needed at a given step.

Since the proposed technique is very customizable and applications to several different topics are presented in this thesis, a direct comparison with other non-structural algorithms is not possible in a generic fashion. Nevertheless we compared each application-specific adaptation of the framework to structural and non-structural state-of-the art domain specific solutions. In all the comparisons our framework has allowed to obtain better results and in many scenarios requiring a comparable amount of computing time.

Future work

We intend to further develop our approach in several ways. First of all we feel the need for a better formalization of the framework itself, in particular by generalizing it in a coherent fashion that is able to express all the domain-specific adaptation explored in this thesis. Moreover we are very encouraged by the result obtained in the field of 3D reconstruction. Specifically we think that our work on surface registration could be significantly improved by the definition of novel feature detectors and descriptors able to exploit the strengths of our framework. Specifically we are already working on a class of robust detectors that are able to select repeatable and distinctive surface locations. Those detectors will be combined with a novel type of relaxed descriptors that are not very distinctive by itself (i.e. they exhibit low precision) but are highly repeatable (i.e. they allow for a high recall). This is the ideal situation for our framework, since we are not disturbed by a high number of outliers, providing that a minimum number of correct matches are in the candidate set. In addition we think that also camera calibration could benefit from our approach, in particular regarding automatic calibration from motion sequences. We are already working on a robust inlier selection method for improved bundle adjustment.

Bibliography

[11] AGRAWAL, A., ANSARI, N., AND HOU, E. Evolutionary programming for fast and robust point pattern matching. *IEEE International Conference on Neural Networks* (1994), 1777–1782.

[12] ALBARELLI, A., BULÒ, S. R., TORSELLO, A., AND PELILLO, M. Matching as a non-cooperative game. In *ICCV 2009: Proceedings of the 2009 IEEE International Conference on Computer Vision* (2009), IEEE Computer Society.

[13] ALBARELLI, A., PELILLO, M., AND VIVIANI, S. Consensus graph for symmetry plane estimation. In *SSPR/SPR* (2008), pp. 173–181.

[14] ALMOHAMAD, H. A polynomial transform for matching pairs of weighted graphs. *J. Applied Math. Modeling 15* (1991).

[15] ALMOHAMAD, H. A., AND DUFFUAA, S. O. A linear programming approach for the weighted graph matching problem. *IEEE Trans. Pattern Anal. Machine Intell. 15*, 5 (1993), 522–525.

[16] AMBLER, A. P., BARROW, H. G., BROWN, C. M., BURSTALL, R. M., AND POPPLESTONE, R. J. A versatile computer-controlled assembly system. In *IJCAI* (1973), pp. 298–307.

[17] ANDRADE NETO, E. L., KHAN, E., WOODS, J. C., AND GHANBARI, M. Segmentation and tracking for interactive sport scenes using region adjacency graphs, picture trees and prior information. In *Picture Coding Symposium* (2003), pp. 353–358.

[18] ANSARI, N., CHEN, M. H., AND HOU, E. S. H. Point pattern matching by a genetic algorithm. In *IECON 90-16th Annual Conference of IEEE Industrial Electronics Society* (Pacific Grove, CA, Nov, 27-30 1990), vol. 2, IEEE, New York, pp. 1233–1238.

[19] ATALLAH, M. J. On symmetry detection. *IEEE Trans. Computers 34*, 7 (1985), 663–666.

[20] BAR-SHALOM, Y., AND FOREMAN, T. *Tracking and Data Association*. Acad.Pr.Inc, 1988.

[21] BARROW, H. G., AND BURSTALL, R. M. Subgraph isomorphism, matching relational structures and maximal cliques. *Inf. Process. Lett. 4*, 4 (1976), 83–84.

[22] BARTALMIO, M., SAPIRO, G., AND RANDALL, G. Morphing active contours. *IEEE Trans. on PAMI 22*, 7 (2000), 733–737.

[23] BATTITI, R., AND PROTASI, M. Reactive local search for the maximum clique problem. *Algorithmica 29*, 4 (2001), 610–637.

[24] BAUM, L. E., AND EAGON, J. A. An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology. *Bull. Amer. Math. Soc. 73* (1967), 360–363.

[25] BEN-ARIE, J., AND WANG, Z. Pictorial recognition of objects employing affine invariance in the frequency domain. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 20*, 6 (1998), 604–618.

[26] BESL, P. J., AND MCKAY, N. D. A method for registration of 3-d shapes. *IEEE Trans. Pattern Anal. Mach. Intell. 14*, 2 (1992), 239–256.

[27] BESL, P. J., AND MCKAY, N. D. A method for registration of 3-d shapes. *IEEE Trans. Pattern Anal. Machine Intell. 14*, 2 (1992), 239–256.

[28] BIGÜN, J. Pattern recognition in images by symmetries and coordinate transformations. *Computer Vision and Image Understanding 68*, 3 (1997), 290–307.

[29] BOROTSCHNIG, H., SINCLAIR, D., AND PINZ, A. Fuzzy graph tracking. In *5th Int'l Symp. on Intelligent Robotic Systems* (1997).

[30] BRADSKI, G. Real time face and object tracking as a component of a perceptual user interface. In *Proc. of WACV* (1998).

[31] BRADSKI, G., AND KAEHLER, A. *Learning OpenCV: Computer Vision with the OpenCV Library*, 1st ed. O'Reilly Media, Inc., October 2008.

[32] BULÒ, S. R. *A game-theoretic framework for similarity-based data clustering*. Ph.D. Thesis, Dipartimento di Informatica, Universita Ca Foscari di Venezia, Italy, 2009.

[33] BULÒ, S. R., AND BOMZE, I. M. Infection and immunization: a new class of evolutionary game dynamics. *Games and Economics Behaviour* (submitted 2009).

[34] BULÒ, S. R., AND PELILLO, M. A continuous characterization of maximal cliques in k-uniform hypergraphs. In *LION* (2007), pp. 220–233.

[35] CAPEL, D. An effective bail-out test for ransac consensus scoring. In *BMVC* (2005), pp. 629–638.

[36] CARCASSONI, M., AND HANCOCK, E. Spectral correspondence for point pattern matching. *Pattern Recognition 36*, 1 (January 2003), 193–204.

[37] CARMICHAEL, O. T., HUBER, D. F., AND HEBERT, M. Large data sets and confusing scenes in 3-d surface matching and recognition. In *3DIM* (1999), pp. 358–367.

[38] CENTER, C. M., CHUM, O., AND MATAS, J. Randomized ransac with t d,d test. In *Image and Vision Computing* (2002), pp. 448–457.

[39] CHEN, C.-S., HUNG, Y.-P., AND CHENG, J.-B. Ransac-based darces: A new approach to fast automatic registration of partially overlapping range images. *IEEE Trans. Pattern Anal. Mach. Intell. 21*, 11 (1999), 1229–1234.

[40] CHEN, W., YANG, H., SU, X., AND TAN, S. Error caused by sampling in fourier transform profilometry. *Optical Engineering 38*, 6 (1999), 1029–1034.

[41] CHEN, Y., AND MEDIONI, G. Object modeling by registration of multiple range images. In *Robotics and Automation, 1991. Proceedings., 1991 IEEE International Conference on* (1991), pp. 2724–2729 vol.3.

[42] CHOI, I., AND CHIEN, S.-I. A generalized symmetry transform with selective attention capability for specific corner angles. *Signal Processing Letters, IEEE 11*, 2 (Feb. 2004), 255–257.

[43] CHRISTMAS, W. J., KITTLER, J., AND M., P. Structural matching in computer vision using probabilistic relaxation. *IEEE Trans. Pattern Anal. Machine Intell. 17* (1995), 749–764.

[44] CHUA, C. S., AND JARVIS, R. Point signatures: A new representation for 3d object recognition. *International Journal of Computer Vision 25*, 1 (October 1997), 63–85.

[45] CHUM, O., AND MATAS, J. Matching with prosac - progressive sample consensus. In *CVPR 05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR05) - Volume 1* (Washington, DC, USA, 2005), IEEE Computer Society, pp. 220–226.

[46] CHUM, O., AND MATAS, J. Optimal randomized ransac. *IEEE Trans. Pattern Anal. Mach. Intell. 30*, 8 (2008), 1472–1482.

[47] CHUM, O., MATAS, J., AND KITTLER, J. Locally optimized ransac. In *DAGM-Symposium* (2003), pp. 236–243.

[48] CHUNG, D. H., YUN, I. D., AND LEE, S. U. Registration of multiple-range views using the reverse-calibration technique. *Pattern Recognition 31*, 4 (1998), 457–464.

[49] COMANICIU, D., RAMESH, V., AND ANDMEER, P. Kernel-based object tracking. *IEEE Trans. on PAMI 25* (2003), 564–575.

[50] CONTE, D., FOGGIA, P., JOLION, J.-M., AND VENTO, M. A graph-based, multi-resolution algorithm for tracking objects in presence of occlusions. *Pattern Recognition 39*, 4 (2006).

[51] COX, G. S., AND DE JAGER, G. A survey of point pattern matching techniques and a new approach to point pattern recognition. In *Proc. South African Symposium on Communications and Signal Processing* (1993), pp. 243–248.

[52] CROSS, A., AND HANCOCK. Graph matching with a dual-step em algorithm. *IEEE Trans. Pattern Anal. Machine Intell. 20* (1998), 1236–1253.

[53] CURLESS, B., AND LEVOY, M. A volumetric method for building complex models from range images. In *SIGGRAPH '96: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques* (New York, NY, USA, 1996), ACM, pp. 303–312.

[54] DUC, B., BIGUN, E., BIGUN, J., MAITRE, G., AND FISHER, S. Fusion of audio and video information for multi modal person authentication. *Pattern Recognition Letters 18* (1997), 835–843.

[55] EDMONDS, J. Paths, trees, and flowers. *Canad. J. Math. 17* (1965), 449–467.

[56] FARISELLI, P., AND CASADIO, R. Prediction of disulfide connectivity in proteins. *Bioinformatics 17* (2001), 957–964.

[57] FELZENSZWALB, P., AND HUTTENLOCHER, D. Pictorial structures for object recognition. *Int. J. Comput. Vision 61* (2005), 55–79.

[58] FELZENSZWALB, P. F., AND HUTTENLOCHER, D. P. Efficient graph-based image segmentation. *Int. J. Comput. Vision 59*, 2 (2004).

[59] FINCH, A., WILSON, R. C., AND R., H. E. Symbolic graph matching with the em algorithm. *Pattern Recognition 31* (1998), 1777–1790.

[60] FISCHLER, M. A., AND BOLLES, R. C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM 24*, 6 (1981), 381–395.

[61] GESÚ, V. D., VALENTI, C., AND STRINATI, L. Local operators to detect regions of interest. *Pattern Recogn. Lett. 18*, 11-13 (1997), 1077–1081.

[62] GEUSEBROEK, J.-M., BURGHOUTS, G. J., AND SMEULDERS, A. W. M. The amsterdam library of object images. *Int. J. Comput. Vision 61*, 1 (2005), 103–112.

[63] GLOVER, F. Genetic algorithms and tabu search - hybrids for optimisation. *Discrete Appl. Math. 49* (1995), 111–134.

[64] GLOVER, F. Tabu search for nonlinear and parametric optimisation (with links to genetic algorithms). *Discrete Appl. Math. 49* (1995), 231–255.

[65] GLOVER, F. Ejection chains, reference structures and alternating path methods for traveling salesman problems. *Discrete Appl. Math. 65* (1996), 223–253.

[66] GOLD, S., AND RANGARAJAN, A. A graduated assignment algorithm for graph matching. *IEEE Trans. Pattern Anal. Machine Intell. 18* (1996), 377–388.

[67] GOMILA, C., AND MEYER, F. Graph-based object tracking. *Proc. ICIP 2* (2003), 41–44.

[68] GOSHTASBY, A. Registration of images with geometric distortions. *IEEE Trans. Geosc. and Rem. Sens.*, 26 (1988), 60–64.

[69] GRACIANO, A., CESAR-JR, R., AND BLOCH, I. Graph-based object tracking using structural pattern recognition. *Proc. of SIBGRAPI* (2007), 179–186.

[70] GRANGER, S., PENNEC, X., AND ROCHE, A. Rigid point-surface registration using an em variant of icp for computer guided oral implantology. In *MICCAI '01: Proceedings of the 4th International Conference on Medical Image Computing and Computer-Assisted Intervention* (London, UK, 2001), Springer-Verlag, pp. 752–761.

[71] GRIFFIN, G., HOLUB, A., AND PERONA, P. Caltech-256 object category dataset. Tech. Rep. 7694, California Institute of Technology, 2007.

[72] GRIFFIN, P., AND ALEXOPOULOS, C. Point pattern matching using centroid bounding. *T-SMC 19* (1989), 1274–1276.

[73] GUAN, C., HASSEBROOK, L., AND LAU, D. Composite structured light pattern for three-dimensional video. *Opt. Express 11*, 5 (2003), 406–417.

[74] GUSFIELD, D. Three fast algorithms for four problems in stable marriage. *SIAM J. Comput. 16*, 1 (1987), 111–128.

[75] GUSFIELD, D., AND IRVING, R. W. *The stable marriage problem: structure and algorithms*. MIT Press, Cambridge, MA, USA, 1989.

[76] HANCOCK, E., AND KITTLER, J. Edge-labeling using dictionary-based relaxation. *IEEE Trans. Pattern Anal. Machine Intell. 12* (1990), 165–181.

[77] HARRIS, C., AND STEPHENS, M. A combined corner and edge detector. In *Proc. Fourth Alvey Vision Conference* (1988), pp. 147–151.

[78] HEDAU, V., ARORA, H., AND AHUJA, N. Matching images under unstable segmentation. In *IEEE Conf. Computer Vision and Patt. Recogn.* (2008).

[79] HERPERS, R., AND SOMMER, G. iscrimination of facial regions based on dynamic grids of point representations. *International Journal of Pattern Recognition and Artificial Intelligence 12* (1998), 381–405.

[80] HORN, B. K. P. Closed-form solution of absolute orientation using unit quaternions. *Journal of the Optical Society of America. A 4*, 4 (Apr 1987), 629–642.

[81] HSU CHANG, S., HSUAN CHENG, T. F., HSING HSU T, W., AND ZUA WU T, G. Fast algorithm for point pattern matching: Invariant to translations rotations and scale changes. *Pattern Recognition* (1997), 311–320.

[82] HUTTENLOCHER, D. P., KLANDERMAN, G. A., AND RUCKLIDGE, W. A. Comparing images using the hausdorff distance. *IEEE Trans. Pattern Anal. Mach. Intell. 15*, 9 (1993), 850–863.

[83] IOFFE, S., AND FORSYTH, D. Human tracking with mixtures of trees. *Computer Vision, IEEE International Conference on 1* (2001), 690.

[84] ISARD, M. Pampas: real-valued graphical models for computer vision. *IEEE Conf. Computer Vision and Patt. Recogn. 1* (2003), 613–620.

[85] ISARD, M., AND BLAKE, A. Condensation - conditional density propagation for visual tracking. *Int. J. Comput. Vision 29*, 1 (1998), 5–28.

[86] JANG, D.-S., JANG, S.-W., AND CHOI, H.-I. 2d human body tracking with structural kalman filter. *Pattern Recognition 35*, 10 (2002), 2041–2049.

[87] JOAN BATLLE, E. M. M., AND SALVI, J. Recent progress in coded structured light as a technique to solve the correspondence problem: a survey. *Pattern Recognition 31*, 7 (1998), 963–982.

[88] JOAQUIM SALVI, J. P., AND BATLLE, J. Pattern codification strategies in structured light systems. *Pattern Recognition 37* (2004), 827–849.

[89] JOHNSON, A. E., AND HEBERT, M. Using spin images for efficient object recognition in cluttered 3d scenes. *IEEE Trans. Pattern Anal. Mach. Intell. 21*, 5 (1999), 433–449.

[90] KADYROV, A. Affine parameter estimation from the trace transform. *IEEE Trans. Pattern Anal. Mach. Intell. 28*, 10 (2006), 1631–1645. Senior Member-Petrou, Maria.

[91] KENNEDY, J., AND EBERHART, R. Particle swarm optimization. *Neural Networks, 1995. Proceedings., IEEE International Conference on 4* (2002), 1942–1948.

[92] KIM, H., AND KIM, J. Hierarchical random graph representation of handwritten characters and its application to hangul recognition. *Pattern Recognition 34* (2001), 187–201.

[93] KIM, J., KOLMOGOROV, V., AND ZABIH, R. Visual correspondence using energy minimization and mutual information. In *Proc. of IEEE Intl Conference on Computer Vision* (2003), pp. 1033–1040.

[94] KIMIA, B. B., TANNENBAUM, A. R., AND ZUCKER, S. W. Shapes, shocks, and deformations i: The components of two-dimensional shape and the reaction-diffusion space. *International Journal of Computer Vision 15* (1994), 189–224.

[95] KITTLER, J., CHRISTMAS, W., AND PETROU, M. Probabilistic relaxation for matching problems in computer vision. In *IEEE Proceedings of the International Conference on Computer Vision (ICCV93)* (1993), pp. 666–673.

[96] KRISHNAMURTHY, V., AND LEVOY, M. Fitting smooth surfaces to dense polygon meshes. In *Proceedings of SIGGRAPH 96* (1996), pp. 313–324.

[97] KUHN, H. W. The hungarian method for the assignment problem. *Naval Research Logistic Quarterly 2* (1955), 83–97.

[98] KULKARNI, P., AND DUTTA, D. An investigation of techniques for asymmetry rectification. *Journal of Mechanical Design 117*, 4 (1995), 620–626.

[99] LEVI, G. A note on the derivation of maximal common subgraphs of two directed or undirected graphs. *Calcolo 9* (1972), 341–354.

[100] LILIENBLUM, E., AND MICHAELIS, B. Optical 3d surface reconstruction by a multi-period phase shift method. *JCP 2*, 2 (2007), 73–83.

[101] LIU, Y. Replicator dynamics in the iterative process for accurate range image matching. *Int. J. Comput. Vision 83*, 1 (2009), 30–56.

[102] LOWE, D. Distinctive image features from scale-invariant keypoints. In *International Journal of Computer Vision* (2003), vol. 20, pp. 91–110.

[103] LOY, G., AND ZELINSKY, A. Fast radial symmetry for detecting points of interest. *IEEE Trans. Pattern Anal. Mach. Intell. 25*, 8 (2003), 959–973.

[104] LU, C., ZHANG, C., WEN, F., AND YAN, P. Principle component analysis-based symmetry detection. *Tien Tzu Hsueh Pao/Acta Electronica Sinica 27*, 5 (1999), 25–28. Cited By (since 1996): 4.

[105] LUCCHESE, L. A frequency domain technique based on energy radial projections for robust estimation of global 2d affine transformations. *Computer Vision and Image Understanding 82*, 1 (2001), 82–83.

[106] LUO, B., AND HANCOCK, E. A robust eigendecomposition framework for inexact graph matching. In *Proceedings 11th International Conference on Image Analysis and Processing* (2001), pp. 465–470.

[107] LUO, B., AND HANCOCK, E. Structural graph matching using the em algorithm and singular value decomposition. *IEEE Transactions on Pattern and Machine Intelligence 23* (2001), 1120–1136.

[108] LUO, B., AND HANCOCK, E. R. A unified framework for alignment and correspondence. *Computer Vision and Image Understanding 92*, 1 (2003), 26–55.

[109] MARDIA, K., AND JUPP, P. *Directional Statistics*. Wiley, 2000.

[110] MARIANI, R. Face learning using a sequence of images. *International Journal of Pattern Recognition and Artificial Intelligence 14* (2000), 631–648.

[111] MASUDA, T. Registration and integration of multiple range images by matching signed distance fields for object shape modeling. *Comput. Vis. Image Underst. 87*, 1-3 (2002), 51–65.

[112] MATAS, J., AND CHUM, O. Randomized ransac with sequential probability ratio test. In *ICCV '05: Proceedings of the Tenth IEEE International Conference on Computer Vision* (Washington, DC, USA, 2005), IEEE Computer Society, pp. 1727–1732.

[113] MCGREGOR, J. Backtrack search algorithms and the maximal common subgraph problem. *Software-Practice and Experience 12* (1982), 23–34.

[114] MEMBER-ZHENG, Y., AND MEMBER-DOERMANN, D. Robust point matching for nonrigid shapes by preserving local neighborhood structures. *IEEE Trans. Pattern Anal. Mach. Intell. 28*, 4 (2006), 643.

[115] MIKOLAJCZYK, K., AND SCHMID, C. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis & Machine Intelligence 27*, 10 (2005), 1615–1630.

[116] MIKOLAJCZYK, K., TUYTELAARS, T., SCHMID, C., ZISSERMAN, A., MATAS, J., SCHAFFALITZKY, F., KADIR, T., AND VAN GOOL, L. A comparison of affine region detectors. *Int. J. Comput. Vision 65*, 1-2 (2005), 43–72.

[117] MOTZKIN, T. S., AND STRAUS, E. G. Maxima for graphs and a new proof of a theorem of Turán. *Canad. J. Math. 17* (1965), 533–540.

[118] MOUNT, D. M., NETANYAHU, N. S., AND MOIGNE, J. L. Efficient algorithms for robust feature matching. *Pattern Recognition 32* (1998), 17–38.

[119] NUMMIARO, K., KOLLER-MEIER, E., AND VAN GOOL, L. An adaptive color-based particle filter. *Image and Vision Computing 21*, 1 (2003), 99–110.

[120] OGAWA, H. Labeled point pattern matching by fuzzy relaxation. *Pattern Recognition 17*, 5 (1984), 569–573.

[121] OLSON, C. F., AND HUTTENLOCHER, D. P. Automatic target recognition by matching oriented edge pixels. *IEEE Transactions on Image Processing 6* (1997), 103–113.

[122] PAPADIMITRIOU, C., AND STEIGLITZ, K. *Combinatorial Optimization*. Prentice-Hall, 1982.

[123] PAVAN, M., AND PELILLO, M. Dominant sets and pairwise clustering. *IEEE Trans. Pattern Anal. Mach. Intell. 29*, 1 (2007), 167–172.

[124] PEARCE, A., CAELLI, T., AND BISCHOF, F. Rulegraphs for graph matching in pattern recognition. *Pattern Recognition 27* (1994), 1231–1247.

[125] PELILLO, M. Replicator equations, maximal cliques, and graph isomorphism. *Neural Computation 11*, 8 (1999), 1933–1955.

[126] PELILLO, M. Replicator equations, maximal cliques, and graph isomorphism. *Neural Comput. 11*, 8 (1999), 1933–1955.

[127] PELILLO, M., SIDDIQI, K., AND ZUCKER, S. W. Attributed tree matching and maximum weight cliques. In *ICIAP '99: Proceedings of the 10th International Conference on Image Analysis and Processing* (Washington, DC, USA, 1999), IEEE Computer Society, p. 1154.

[128] PELILLO, M., SIDDIQI, K., AND ZUCKER, S. W. Matching hierarchical structures using association graphs. *IEEE Trans. Pattern Anal. Machine Intell. 21*, 11 (1999), 1105–1120.

[129] PODOLAK, J., SHILANE, P., GOLOVINSKIY, A., RUSINKIEWICZ, S., AND FUNKHOUSER, T. A planar-reflective symmetry transform for 3d shapes. In *SIGGRAPH '06: ACM SIGGRAPH 2006 Papers* (New York, NY, USA, 2006), ACM, pp. 549–559.

[130] PODOLAK, J., SHILANE, P., GOLOVINSKIY, A., RUSINKIEWICZ, S., AND FUNKHOUSER, T. A planar-reflective symmetry transform for 3d shapes. *ACM Trans. Graph. 25*, 3 (2006), 549–559.

[131] RAMANAN, D., FORSYTH, D., AND ZISSERMAN, A. Tracking people by learning their appearance. *IEEE Trans. on PAMI 29*, 1 (Jan. 2007), 65–81.

[132] RANADE, S., AND ROSENFELD, A. Point pattern matching by relaxation. *Pattern Recognition 12*, 4 (1980), 269–275.

[133] REISFELD, D., WOLFSON, H., AND YESHURUN, Y. Context free attentional operators: the generalized symmetry transform. *Int. J. of Computer Vision, Special Issue on Qualitative Vision* (1994).

[134] ROLLAND, E., PIRKUL, H., AND GLOVER, F. Tabu search for graph partitioning. *Ann. Oper. Res. 63* (1996), 200–232.

[135] ROTA BULÒ, S., TORSELLO, A., AND PELILLO, M. A continuous-based approach for partial clique enumeration. In *GbRPR* (2007), pp. 61–70.

[136] RUSINKIEWICZ, S., AND LEVOY, M. Efficient variants of the icp algorithm. In *Proceedings of the Third Intl. Conf. on 3D Digital Imaging and Modeling* (2001), pp. 145–152.

[137] RUSTAMOV, R. M. Augmented symmetry transforms. In *SMI '07: Proceedings of the IEEE International Conference on Shape Modeling and Applications 2007* (Washington, DC, USA, 2007), IEEE Computer Society, pp. 13–20.

[138] SALVI, J., MATABOSCH, C., FOFI, D., AND FOREST, J. A review of recent range image registration methods with accuracy evaluation. *Image Vision Comput. 25*, 5 (2007), 578–596.

[139] SCHARSTEIN, D., AND SZELISKI, R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithm. *Int. J. of Comp. Vision 47*, 1 (2002), 7–42.

[140] SCHINDLER, G., AND DELLAERT, F. A rao-blackwellized parts-constellation tracker. In *WDV* (2006), pp. 178–189.

[141] SEBASTIAN, T. B., KLEIN, P. N., AND KIMIA, B. B. Recognition of shapes by editing their shock graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence 26*, 5 (2004), 550–571.

[142] SHAH, M. I., AND SORENSEN, D. C. A symmetry preserving singular value decomposition. *SIAM J. Matrix Anal. Appl. 28*, 3 (2006), 749–769.

[143] SHAPIRO, L., AND HARALICK, R. Organization of relational models for scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell. 4*, 6 (November 1982), 595–602.

[144] SHEN, D., IP, H. H.-S., CHEUNG, K. K. T., AND TEOH, E. K. Symmetry detection by generalized complex (GC) moments: A close-form solution. *IEEE Transactions on Pattern Analysis and Machine Intelligence 21*, 5 (1999), 466–476.

[145] SHOKOUFANDEH, A., DICKINSON, S. J., SIDDIQI, K., AND ZUCKER, S. W. Indexing using a spectral encoding of topological structure. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on 2* (1999), 2491.

[146] SIGAL, L., BHATIA, S., ROTH, S., BLACK, M., AND ISARD, M. Tracking loose-limbed people. *IEEE Conf. Computer Vision and Patt. Recogn. 1* (2004), 421–428.

[147] SRINIVASAN, V., LIU, H. C., AND HALIOUA, M. Automated phase-measuring profilometry: a phase mapping approach. *Appl. Opt. 24*, 2 (1985), 185–188.

[148] STARINK, J., AND BACKER, E. Finding point correspondences using simulated annealing. *Pattern Recognition 28*, 2 (1995), 231–240.

[149] SU, X., AND CHEN, W. Fourier transform profilometry: a review. *Optics and Lasers in Engineering 35* (May 2001), 263–284.

[150] SUDDERTH, E., IHLER, A., FREEMAN, W., AND WILLSKY, A. Nonparametric belief propagation. *IEEE Conf. Computer Vision and Patt. Recogn. 1* (2003), 605–612.

[151] SUN, C. Symmetry detection using gradient information. *Pattern Recogn. Lett. 16*, 9 (1995), 987–996.

[152] SUN, C., AND SHERRAH, J. 3d symmetry detection using the extended gaussian image. *IEEE Trans. Pattern Anal. Mach. Intell. 19*, 2 (1997), 164–168.

[153] SURREL, Y. Design of algorithms for phase measurements by the use of phase stepping. *Appl. Opt. 35*, 1 (1996), 51–60.

[154] TAKEDA, M., AND MUTOH, K. Fourier transform profilometry for the automatic measurement of 3-d object shapes. *Appl. Opt. 22*, 24 (1983), 3977–3982.

[155] TANG, F., AND TAO, H. Object tracking with dynamic feature graph. *2nd Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance* (15-16 Oct. 2005), 25–32.

[156] TAREL, J.-P., CIVI, H., AND COOPER, D. B. Pose estimation of free-form 3d objects without point matching using algebraic surface models. In *Proceedings of IEEE Workshop Model Based 3D Image Analysis* (Mumbai, India, 1998), pp. 13–21.

[157] TON, J., AND JAIN, A. Registering landsat images by point matching. *IEEE Trans. Geoscience and Remote Sensing 27*, 5 (September 1989), 642–651.

[158] TORDOFF, B., AND MURRAY, D. W. Guided sampling and consensus for motion estimation. In *ECCV '02: Proceedings of the 7th European Conference on Computer Vision-Part I* (London, UK, 2002), Springer-Verlag, pp. 82–98.

[159] TORR, P. H. S., AND ZISSERMAN, A. Mlesac: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding 78* (2000), 2000.

[160] TORSELLO, A., ALBARELLI, A., AND PELILLO, M. Matching relational structures using the edge-association graph. In *Int. Conf. on Image Analysis and Processing* (Washington, DC, USA, 2007), IEEE Computer Society, pp. 775–780.

[161] TORSELLO, A., BULÒ, S. R., AND PELILLO, M. Grouping with asymmetric affinities: A game-theoretic perspective. In *CVPR '06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Washington, DC, USA, 2006), IEEE Computer Society, pp. 292–299.

[162] TORSELLO, A., AND HANCOCK, E. R. Computing approximate tree edit distance using relaxation labeling. *Pattern Recognition Letters 24* (2003), 1089–1097.

[163] TORSELLO, A., AND HIDOVIC-ROWE, D. Polynomial-time metrics for attributed trees. *IEEE Trans. Pattern Anal. Mach. Intell. 27*, 7 (2005), 1087–1099. Senior Member-Pelillo, Marcello.

[164] TORSELLO, A., ROTA BULÒ, S., AND PELILLO, M. Grouping with asymmetric affinities: a game-theoretic perspective. In *IEEE Conf. Computer Vision and Patt. Recogn.* (2006), pp. 292–299.

[165] TURK, G., AND LEVOY, M. Zippered polygon meshes from range images. In *SIGGRAPH '94: Proc. of the 21st annual conference on Computer graphics and interactive techniques* (New York, NY, USA, 1994), ACM, pp. 311–318.

[166] ULLMANN, J. An algorithm for subgraph isomorfism. *Journal of the Association for Computing Machinery 23* (1976), 31–42.

[167] UMEYAMA, S. An eigendecomposition approach to weighted graph matching problems. *IEEE Trans. Pattern Anal. Machine Intell. 10* (1988), 695–703.

[168] UMEYAMA, S. An eigendecomposition approach to weighted graph matching problems. *IEEE Trans. Pattern Anal. Machine Intell. 10*, 5 (1988), 695–703.

[169] UMEYAMA, S. Least-squares estimation of transformation parameters between two point patterns. *IEEE Trans. Pattern Anal. Mach. Intell. 13*, 4 (1991), 376–380.

[170] UMEYAMA, S. Parameterized point pattern matching and its application to recognition of object families. *IEEE Trans. Pattern Anal. Mach. Intell. 15*, 2 (1993), 136–144.

[171] VEENMAN, C., REINDERS, M., AND BACKER, E. Resolving motion correspondence for densely moving points. *IEEE Trans. on PAMI 23*, 1 (2001), 54–72.

[172] WANG, W.-H., AND CHEN, Y.-C. Point pattern matching by line segments and labels. *Electronics Letters 33*, 6 (1997), 478–479.

[173] WEIBULL, J. *Evolutionary Game Theory*. MIT Press, 1995.

[174] WILLIAMS, M., WILSON, R., AND HANCOCK, E. Deterministic search for relational graph matching. *Pattern Recognition 32* (1999), 1255–1271.

[175] WILSON, R. C., AND R., H. E. Structural matching by discrete relaxation. *IEEE Trans. Pattern Anal. Machine Intell. 19* (1997), 634–648.

[176] WILSON, R. C., AND R., H. E. Graph matching with hierarchical discrete relaxation. *Pattern Recognition Letters 20* (1999), 1041–1052.

[177] WOLTER, J. D., WOO, T. C., AND VOLZ, R. A. Optimal algorithms for symmetry detection in two and three dimensions. *The Visual Computer 1*, 1 (1985), 37–48.

[178] XIAOBO, C., TONG, X. J., TAO, J., AND YE, J. Research and development of an accurate 3d shape measurement system based on fringe projection: Model analysis and performance evaluation. *Precision Engineering 32*, 3 (2008), 215–221.

[179] YANG, R., AND GAO, Y. Line-based affine invariant object location using transformation space decomposition. In *ICPR '06: Proceedings of the 18th International Conference on Pattern Recognition* (Washington, DC, USA, 2006), IEEE Computer Society, pp. 646–649.

[180] YANG, Z., AND COHEN, F. S. Cross-weighted moments and affine invariants for image registration and matching. *IEEE Trans. Pattern Anal. Mach. Intell. 21*, 8 (1999), 804–814.

[181] YILMAZ, A., JAVED, O., AND SHAH, M. Object tracking: A survey. *ACM Journal of Computing Surveys 38*, 4 (2006).

[182] YIN, P.-Y. Particle swarm optimization for point pattern matching. *J. Visual Communication and Image Representation 17*, 1 (2006), 143–162.

[183] YUE, H.-M., SU, X.-Y., AND LIU, Y.-Z. Fourier transform profilometry based on composite structured light pattern. *Optics Laser Technology 39* (Sept. 2007), 1170–1175.

[184] YUEN, P. Dominant point matching algorithm. *Electron. Lett. 29* (1993), 2023–2024.

[185] ZABRODSKY, H., PELEG, S., AND AVNIR, D. Symmetry as a continuous feature. *IEEE Trans. on Pattern Analysis and Machine Intelligence 17*, 12 (1995), 1154–1166.

[186] ZAHARESCU, A., BOYER, E., VARANASI, K., AND HORAUD, R. P. Surface feature detection and description with applications to mesh matching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Miami Beach, Florida, June 2009).

[187] ZHANG, K. A constrained edit distance between unordered labeled trees. *Algorithmica 15*, 6 (1996), 205–222.

[188] ZHANG, L., XU, W., AND CHANG, C. Genetic algorithm for affine point pattern matching. *Pattern Recogn. Lett. 24*, 1-3 (2003), 9–19.

[189] ZHANG, Z. Iterative point matching for registration of free-form curves and surfaces. *Int. J. Comput. Vision 13*, 2 (1994), 119–152.

[190] ZHONG, J., AND ZHANG, Y. Absolute phase-measurement technique based on number theory in multifrequency grating projection profilometry. *Appl. Opt. 40*, 4 (2001), 492–500.