Image Registration by Probabilistic Multi-Assignment Graph Matching

Uri Okun

Image Registration by Probabilistic Multi-Assignment Graph Matching

Research Thesis

As Partial Fulfillment of the Requirements for
the Degree Master of Science in Electrical Engineering

Uri Okun

Submitted to the Senate of the Technion—Israel Institute of Technology

Sivan 5774 Haifa June 2014

Acknowledgement

The Research Thesis Was Done Under The Supervision of Professor Israel Cohen from the Department of Electrical Engineering at the Technion and Professor Yosi Keller from the School of Engineering at Bar-Ilan University.

The Generous Financial Help Of The Technion Is Gratefully Acknowledged.
This research was supported by the Israel Science Foundation (grant no. 1130/11).

I would like to express my gratitude to my supervisors, Prof. Yosi Keller and Prof. Israel Cohen for the supervision, guidance and support throughout this research.
I would also like to thank Dovi Weinberger and Dr. Gili Telem for the helpful comments and the fruitful discussions.
Special thanks to my dear family - my parents, Yael and David, my parents-in-law, Orly and Yossi, and my brothers, Jonathan and Avner, for the continuous encouragement and support.
Last but not least, I would like to express my deep love and gratitude to my beloved wife Hila. Thank you for your patience, your endless love and constant support. Without you none of this would have happened.

Contents

List of Figures

List of Algorithms

Abstract

Image registration is the task of finding a large consistent set of corresponding points or features between two images depicting the same scene. The common general methodology for solving such a problem can be outlined as a two stage process: In the first, features of interest, such as corners or spatially localized blobs, are detected in both images, and in the second, those features are matched using geometrical and appearance related cues.

Two distinct paradigms for conducting the matching stage have evolved in parallel during the last decades: One robustly estimates a parametric transformation or a parametric constraint (e.g. the pinhole perspective epipolar constraint) over the set of correspondences, while pruning non compliant features. The second makes use of spatial relations between small groups of features, in order to find geometric consistency in the transformation from one image to the other, by exploiting graph theory results. Both methodologies often use appearance similarity to initiate their strategies. While the former strategy enables efficient matching when a global motion model is known to exist and the appearance-based initialization suffices, the latter is more robust to the starting point as it uses more than one putative match for each point, and more flexible to nonparametric deformations. Nevertheless, the heuristic assumption of point groups consistency is not guaranteed, especially when matching a wide-baseline stereo pair, causing performance degradation as the perspective deformation increases.

In this thesis, the points matching problem is addressed, with a special concern to the wide-baseline stereo matching problem. We propose a unified probabilistic image matching algorithm, namely MAGMA (Multi-Assignment Graph Matching Algorithm), that utilizes both piecewise graph matching and epipolar-geometry-based constraints. First, we extend the applicability of the graph matching approach to wide-baseline stereo matching, by generalizing it to multiple unconnected components, each of its own geometric consistent transformation. For that we derive a projected iterative least squares optimization scheme that imposes the matching

restrictions in the course of the solution. Second, we propose an approach to fusing independent sources of information with the small groups consistency, to improve the graph matching process robustness, and suggest a probabilistic measure encoding the epipolar constraint for this task. Finally, we combine the two steps in a probabilistic formulation to derive a stable, fast-converging iterative process. The proposed framework is general and can be altered so that the epipolar prior is replaced by any other prior or combination of priors, induced by a-priori information such as appearance similarity or some rigid transformation.

Our approach is comprehensively verified by applying it to simulated data in various challenging setups, as well as real image sequences. MAGMA is shown to compare favorably to contemporary state-of-the-art algorithms, and even outperforms them in many cases.

Notation

ε	nonnegative tolerance parameter - kernel width
θ_{ijk}	angle formed by the three points x_i, x_j and x_k
$\mathbf{\Lambda}$	affine constraint offset matrix
λ	SCMF regularization parameter
π	plane in space
Σ_f	fundamental matrix covariance
Σ_l	epipolar line covariance
σ	image points' standard deviation estimation
$\phi_o(\cdot, \cdot, ...)$	affinity measure of order o
χ^2	Chi-Square probability distribution function
$\mathbf{\Psi}$	affine constraint scale matrix
$\mathbf{\Psi}^{\text{cols}}$	columns normalization constraint scale matrix
$\mathbf{\Psi}^{\text{rows}}$	rows normalization constraint scale matrix
$\mathbf{\Omega}$	soft multi-assignment matrix
$\mathbf{A}$	affinity matrix
C_1, C_2	cameras' positions in world's coordinates
C	assignment (set of correspondences)
$C^{(m)}$	one set of correspondences
$c_{ii'}$	corresponding pair of points, x_i and $y_{i'}$
e, e'	epipole points
$\mathbf{F}$	fundamental matrix
$\mathbf{F}_n$	fundamental matrix estimate at iteration n
$F_{\chi^2_m}(\cdot)$	Chi-Square cumulative probability with m degrees of freedom
f_1, f_2	focal-lengths of cameras 1 and 2 (resp.)

$\mathbf{I}$	identity matrix
I_1, I_2	source and target images (resp.)
k^2	square Mahalanobis distance
k_{nn}	number of nearest neighbors for each point in S_1
$L(\mathbf{p}, \mathbf{F})$	likelihood of assignment probabilities and fundamental matrix
$L_E(\mathbf{p}, \mathbf{F})$	likelihood of the epipolar model
$L_G(\mathbf{p}, \mathbf{F})$	likelihood of probabilistic graph matching
l, l', m, m'	epipolar lines in images' coordinates
M	number of calculated assignments
$MaxIter$	maximum number of iterations in MAGMA
N	number of rows in $\mathbf{\Omega}$
N_b	cardinality of S_b
n	cardinality of C
$P_Q(\cdot)$	orthogonal projection operator to subspace Q
$P(u(c_{ii'}))$	unary probability of a single match
p	probability scalar
$\mathbf{p}$	assignment probabilities vector
$\mathbf{p}_n$	assignment probabilities vector estimate at iteration n
Q	linear constraints subspace
S_b	set of points in image I_b
TH_σ	threshold for inliers' standard deviation in MAGMA
TH_{ratio}	threshold for standard deviation relative change in MAGMA
t	robust estimation error threshold parameter
$\mathbf{W}^{(m)}$	one soft assignment matrix
$\left[\mathbf{W}^{(m)}\right]_{ij}$	the (i, j) entry of the matrix $\mathbf{W}^{(m)}$
$\mathbf{w}$	soft assignment row-stacked matrix
$\mathbf{w}^{(m)}$	row-stacked vector of the soft assignment matrix $\mathbf{W}^{(m)}$
X	point in space
x	image point in the source image
y	image point in the target image
$\tilde{x}$,$\tilde{y}$	image points x and y (resp.) in homogeneous coordinates

x_i, y_i	ith point in image I_1, and I_2 (resp.)
$\{x, y\}$	corresponding pair of points, x and y
$\mathbf{Z}$	binary assignment matrix
z_{ij}	the (i, j) entry of the matrix $\mathbf{Z}$
$\mathbf{z}$	row-stacked replica of $\mathbf{Z}$
$\mathbf{0}$	matrix of zeros
$\mathbf{1}$	matrix of ones
$\|\cdot\|_{L_2}$	Euclidian norm
$(\cdot)^*$	estimator (or an optimal value)
$[\cdot]_+$	nonnegative projection operator

Abbreviations

ALS	Alternating Least Squares
BEEM	Balanced Exploration and Exploitation Model
BLOGS	Balanced LOcal and Global Search
DoG	Difference of Gaussians
GA	Graduated Assignment
GM	Graph Matching
HGM	HyperGraph Matching
IPFP	Integer Projected Fixed Point
JFD	Joint Feature Distribution
LIDAR	LIght Detection And Ranging
LO-RANSAC	Locally Optimized RANdom SAmpling Consensus
MAGMA	Multi-Assignment Graph Matching Algorithm
MAPSAC	Maximum A-Posteriori SAmpling Consensus
MLESAC	Maximum Likelihood Estimation SAmpling Consensus
NAPSAC	N-Adjacent Points SAmpling Consensus
NMF	Nonnegative Matrix Factorization
PDF	Probability Density Function
PROSAC	PROgressive SAmpling Consensus
QBP	Quadratic Binary Problem
RANSAC	RANdom SAmpling Consensus
RBF	Radial Basis Function
RMA	Rank-M-Approximation
ROA	Rank-One-Approximation
ROC	Receiver Operating Characteristic

RRWM	Reweighted Random Walk Matching
SCMF	Stochastic Constrained Matrix Factorization
SIFT	Scale Invariant Feature Transform
SM	Spectral Matching
SSA	Spectral Symmetry Analysis
STD	STandard Deviation
TGM	Tensor Graph Matching
THOA	Tensor High Order Assignment

Chapter 1

Introduction

1.1 Background and Motivation

Image registration is a fundamental task in computer vision, whose applications include, inter alia, object recognition [1], three-dimensional scene reconstruction, motion estimation, aerial photography change detection, data fusion and mapping applications.

Contemporary image registration algorithms often use interest points or local features for matching. It is common for such algorithms to consist of the following three steps: First, the *detection* of interest points or image features, such as corners [2], local maxima/minima of DoG operator in scale-space [3], or stable extremal regions of gray-levels [4]. Second, the computation of *descriptors* - the detected feature points are each encoded by a descriptor vector that discriminates between non-corresponding pairs of points, and is invariant to certain differences in acquisition conditions, such as lightning conditions, scale, rotations, and affine transformations. Last, *matching* the features by their descriptors' similarity.

The matching step (also known as *assignment* step) is considered to be a tough problem, due to inter-image variability that stems from the difference in cameras' pose and the changes in the observed scene. A naive implementation of the matching step would be by assigning each point in the source image to its nearest neighbor in the target, in terms of the distance between their corresponding descriptors. However, this method often yields poor results, and a lot of wrong matches are formed. Some causes for that are: false matches between repetitive shapes or objects, occlusions or non-overlapping regions (in which case the feature in one image does not exist in the second one), or accidental feature similarity.

The matching of a wide-baseline stereo pair, where the two images are taken from disparate positions and viewing angles, is of particular difficulty and interest, due to significant perspective deformations, making the matching difficult. As the baseline angle grows the more difficult it is to compensate for the appearance dissimilarity between corresponding feature points, thus making the descriptors labeling more difficult, and causing for degradation in inlier rate. Moreover, As the perspective deformation increases the global motion model becomes more complex, making it even harder to recover the true correspondences (an example of such a difficult case is hereby presented in Fig. 1.1).

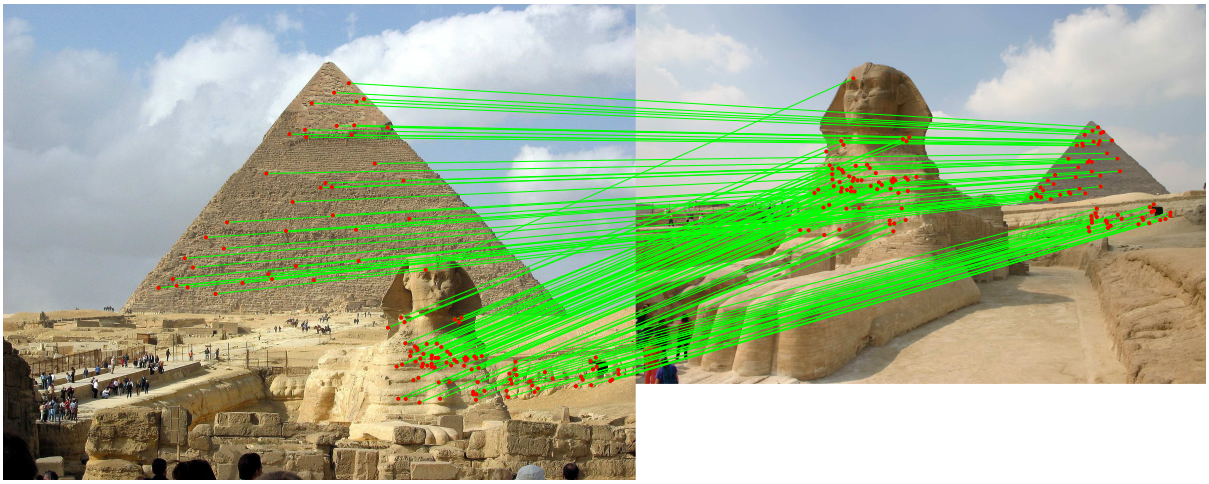


Figure 1.1: Wide-baseline image registration. Both local and global perspective deformations are present in the scene, making it harder to find a proper matching. The registration presented was achieved by the algorithm proposed in this paper.

Over the past two decades the computer vision community put a tremendous effort to solve the assignment problem, using both visual and spatial cues. Two of the main approaches for solving this problem are based upon geometrical relations between distinct feature points in both images. One common approach is to first use the descriptors metrics in order to construct a list of putative matches based on their appearance similarity, and then filter the putative matches using geometric constraints such as the epipolar constraint [5, 6]. In recent years there has been a progress in the research of robust estimation and filtering techniques for this purpose [7–13], pushing the limit of robust geometry estimation in the case of high outlier percentage. However, this approach is not without its drawbacks especially in the case of repetitive patterns in the direction parallel to the camera's motion, and in high outliers rates scenarios. Another modern

approach, based on graph theory, is to search for a consistent set of correspondences between the two sets of features, according to their local or global geometric-structural relations. A myriad of algorithms were developed in the last few years for second order (pairwise) [14–17] or higher order (mostly triplets) [18–20] matching. Graph matching (GM) schemes proved to be efficient and robust to both noise and outliers. Also, they enable considering more than just one target feature for each point in the source image, as opposed to the classical robust estimation and filtering techniques.

The main drawback of GM approaches is the lack of flexibility in the motion model. Namely, such approaches usually enforce simple, relaxed but rigid transformations such as isometry or similarity, and so cannot cope with more complex real-world transformations. The pairwise matching scheme is suitable for approximating isometry transformation (translation and rotation). The more complicated triplets GM cope well with scale changes (similarity transformations), and even affine transformations or plane perspective (homographic) transformations could be dealt using 4th and 5th order matching respectively (those are usually considered infeasible due to the exponential growth in complexity along with graph order). Nevertheless, even the highest order GM schemes still pose a restriction for approximately rigid and global transformations.

In the case of wide-baseline stereo matching, both matching strategies suffer from performance degradation: On one hand, epipolar robust estimation techniques start failing when the inlier rate within the putative set of correspondences becomes insufficient for the formation of a strong enough consensus set to be used for model estimation. On the other hand, GM algorithms (of any order) are inadequate to deal with complex or discontinuous transformations, which are often formed when the baseline is large relative to the scene depth variation.

1.2 Overview

In this thesis we address the stereo matching problem. We focus on matching a wide-baseline image pair, though some of the proposed methods and concepts could also facilitate other matching scenarios, such as faces or non-rigid objects matching.

When matching a wide-baseline stereo image pair, the transformation can take any form compliant with the aforementioned epipolar constraint. A perspective transformation is in gen-

eral non rigid and non parametric, as it depends on cameras' parameters and on the scene's depth, which can be non smooth or even discontinuous. However, in real-world situation, it is not completely the case as most scenes are composed of several unconnected objects, each of a roughly constant depth from camera, and each yields a different approximately rigid transformation (see some examples from the BEEM dataset in Fig. 1.2, or different views of the Versailles palace in France in Fig. 1.3).

(a)

(b)

Figure 1.2: Two examples from the BEEM dataset, in which there are several objects in the scene, each undergoes a different transformation between the images. (a) Flowerpot scene. (b) Board scene.

This observation motivates a multiple solution graph matching approach, and gives rise to our proposed Multi-Assignment Graph Matching Algorithm (MAGMA). The suggested approach takes advantage of the GM strength of rigid robust cluster matching, without giving up

Figure 1.3: Several figures of Versailles palace from different angles. This scene demonstrates the need for disconnected subgraphs assignment in natural scenes.

the flexibility of complex perspective transformation. This is done by first relaxing the demand for one global consistent component and replacing it with the more suitable model of several local components, thus, reformulating the problem as a graph clustering (or multi-assignment graph matching) problem. Second, the epipolar geometry inferred from the putative matches is used as a global constraint binding these assignments together, and improving the robustness to outliers and noise. Finally, the two steps are combined in a probabilistic fashion to form a stable, fast-converging iterative process.

The suggested approach takes advantage of both the strengths of epipolar robust estimation and GM methods, while mitigating their shortcomings. Unlike epipolar robust estimation algorithms, the proposed approach exploits the spatial redundancies in the scene (i.e. the heuristic assumption of several consistent and approximately rigid transformations), and unlike GM algorithms, it better describes the motion model, without presuming oversimplified assumption of an approximately rigid motion model. Moreover, the robust estimation initialization of the

proposed approach is of higher inlier ratio than in existing methods, since it is based upon the generalized GM's output. Later on, the graph is better capturing the true matching probabilities, since it incorporates additional, substantial measure of compliance with perspective stereo transformation, based on the so-far obtained data.

In this paper we present the following contributions:

1. We propose replacing the common one-component (or one-assignment) graph matching approach with a multi-assignment one, which better suits the wide-baseline setup.

2. We derive a new optimization scheme for solving the multiple assignment problem in a probabilistic way, via factorizing the affinity graph using a stochastic semi-normalized constrained factorization (SCMF), which can be seen as a special case of nonnegative matrix factorization (NMF).

3. We propose a rigorous algorithm for probabilistically fusing the epipolar geometry constraint and the suggested multi-assignment graph matching approach. The unified framework, maximizes the likelihood of both the structure similarity, and the estimated epipolar geometry. This framework is general, so that the epipolar prior could be replaced by any other probabilistic prior or combination of priors, such as geometrical priors, or appearance-similarity based priors.

The proposed approach is tested and verified by a series of experiments, including both synthetic challenging scenarios and real image sequences, that constitute a large scale quantitative evaluation. The algorithm is tested versus various algorithms from both schools of robust epipolar geometry estimation and graph matching, and turns out to compare favorably to (and mostly outperform) contemporary state-of-the-art matching approaches, showing augmented resiliency to noise, outliers, scale changes, and most of all perspective deformations.

1.3 Thesis Structure

The rest of this thesis is organized as follows: In Chapter 2 we survey relevant algorithms in both graph matching, and epipolar geometry estimation together with their probabilistic interpretations. We discusses the existing methods' strengths and weaknesses. Chapter 3 suggests a

novel approach for solving the matching problem in the wide-baseline setup. Different aspects of the proposed method are presented and the full algorithm is outlined. Some experimental results are shown and discussed in Chapter 4. Concluding remarks and further research directions are discussed in Chapter 5.

Chapter 2

Review of Related Work

2.1 Introduction

In this chapter we review the two major schools of image matching: robust estimation of epipolar geometry and graph matching techniques. Both strategies attempt to find consistent sets of correspondences by incorporating spatial information in the matching process. Section 2.2 presents the notion of epipolar geometry between a stereo image pair, and provides a survey of recent algorithms for robustly estimating it in the presence of outliers. A short survey of epipolar probabilistic characteristics and measures is also provided. Section 2.3 overviews some state-of-the-art algorithms for image registration using graph matching approaches, and some related aspects of the problem at hand.

2.2 Epipolar Geometry Estimation

Given a stereo pair, consists of two images, I_1 and I_2 depicting the same scene, each point correspondence must obey the underlying epipolar constraint. The epipolar constraint is the geometric structure which holds for all pairs of pinhole cameras, regardless of the captured scene, and as such can be used to improve points matching. It depends solely on the cameras' parameters and relative pose. Both the meaning and derivation of the epipolar geometry are thoroughly reviewed in the literature (see for example the work of Hartley and Zisserman [6] and Zhang, and Zhengyou [5]), however a brief review is presented here.

2.2.1 Epipolar Geometry and the Fundamental Matrix

Each image point $x \in I_1$ determines a ray in 3D space, that is in turn back-projected to a line, $l' \in I_2$ as shown in Fig. 2.1a. The stereo geometry dictates that the point $x \in I_1$, the corresponding point $y \in I_2$, the two rays associated with each of them, and the two camera centers in space - C_1 and C_2, are all coplanar. This plane π is determined by the baseline (the line connecting C_1 and C_2) and the ray defined by x. Its intersection with the two image planes defines the two corresponding *epipolar lines* in both images l and l', respectively. The locus of the point y, corresponding to $x \in l \in I_1$ is the line $l' \in I_2$. In Fig. 2.1b we show the family of planes defined by different points X in space. This family, known as an *epipolar pencil*, produces a set of epipolar lines. All epipolar lines in each image intersect in a single point, called the *epipole*.

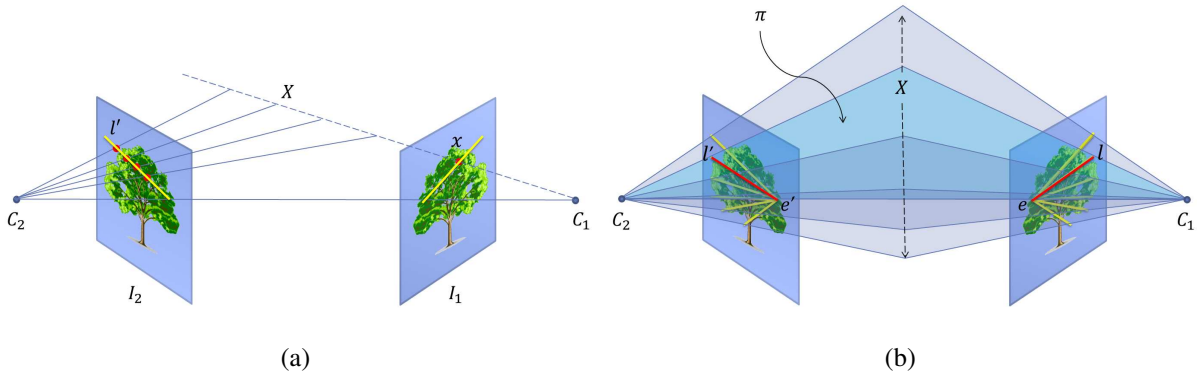


Figure 2.1: Two cameras configuration example. (a) The image point $x \in I_1$ and the first camera center C_1 define a ray in 3D-space. Each world point X on the ray is back projected (through C_2) to a point $y \in I_2$. the locus of all such points in I_2 is a line $l' \in I_2$. (b) The cameras' centers C_1 and C_2, the 3D-space point X, and its images $x \in I_1$ and $y \in I_2$ all lie in a common plane π. Any plane containing the baseline CC' is an epipolar plane. The set of all such planes is known as the epipolar pencil. The baseline intersects the image planes of I_1 and I_2, at the epipoles e and e', respectively, and each plane π intersects the image planes in corresponding epipolar lines l and l'.

Algebraically, this structure is defined by the following equation called the *Epipolar Constraint*:

$$\tilde{x}\mathbf{F}\tilde{y} = 0, \tag{2.1}$$

where $\tilde{x}$ and $\tilde{y}$ are the points x and y in homogeneous coordinates, and $\mathbf{F}$ is a 3×3 rank 2 matrix

known as the *Fundamental matrix*. The Fundamental matrix is independent of scene structure, and depends solely on the cameras' internal parameters and relative pose. The epipolar lines are given by:

$$l' = \tilde{x}\mathbf{F},\ l = \mathbf{F}\tilde{y} \tag{2.2}$$

The fundamental matrix can be calculated using a non-linear seven-point algorithm (e.g. by minimizing the sum of point-to-epipolar-line square distances [6]), or the linear normalized eight-point algorithm [21]. The epipolar geometry is an essential tool for image matching, as it enables limiting the search region for each correspondence to the vicinity of the epipolar line, or alternatively allows for the pruning of outlier correspondences.

2.2.2 Epipolar Robust Estimation for Image Matching

Many registration schemes relay on estimating the fundamental matrix coefficients from sets of putative matches while simultaneously removing gross outliers. These techniques differ from classical estimation procedures such as least squares, as they are intended for robust fit given that the data is contaminated by both noise and outliers.

Over the past few years there has been a tremendous progress in the field, pushing the limit on estimating model parameters faster and more robust to outliers than ever. One early successful technique is RANSAC (random sampling consensus) [22]. RANSAC is an iterative algorithm in which each iteration consists of two steps: 1. A minimal set is drown from the set of putative matches, and 2. the minimal set instantiates a model, which is next validated against the entire putative set. Eventually the model chosen is the one with the highest score (or largest consensus amongst the initial set of putative corresponding pairs).

Many modern robust model estimation algorithms were inspired by RANSAC, and are using a similar hypothesize-and-verify framework. MLESAC (maximum likelihood estimation sampling consensus) [8] soften the cost function for each correspondence, while formalizing the problem as a maximum likelihood estimation, assuming narrow normal distribution for inliers, and uniform broad distribution for outliers. MAPSAC (maximum a-posteriori estimation sampling consensus) [9] further develops the concept, following a Bayesian approach. NAPSAC (N adjacent points sampling consensus) [10] exploits the fact that true inliers tend to be in a close proximity to one another. It samples each minimal set within a fixed radius from one randomly

sampled point. PROSAC (progressive sampling consensus) [11] uses matching scores, such as SIFT (Scale-Invariant Feature Transform) proximity, or the ratio between the closest to the second closest SIFT descriptor distances, to guide the sampling process, where progressively larger subsets are drown in each iteration in a descending order of their score. LO-RANSAC (locally optimized random sampling consensus) [7] adds a local optimization step whenever a best-so-far model is generated during the RANSAC iterations.

More recent approaches combine guided local sampling with global search. Such are BEEM (balanced exploration and exploitation model search algorithm,) [12] and BLOGS (balanced local and global search) [13] algorithms. BEEM balances between local and global exploration steps, and local exploitation, similar to the one preformed in LO-RANSAC. Both explorations steps are guided by an empirical probability of outliers' and inliers' correspondences, based on their closest to second closest SIFT proximity ratio. The local exploration step is done by giving priority to correspondences sampled from the best-so-far model consensus set. BLOGS is similar in spirit to BEEM. It consists of two steps: A global search ("jump") guided by features' similarity, and a local search ("diffusion") guided by JFD (joint feature distribution) [23] which gives a probabilistic search region for each point correspondence, based on the best-so-far generated sample set. Each iteration, one of these steps is preformed at random.

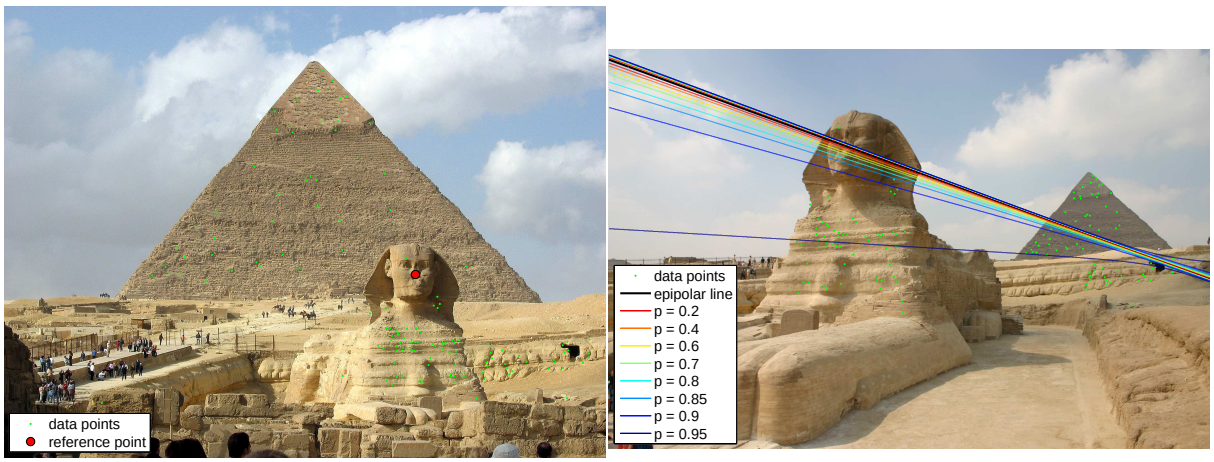
2.2.3 Epipolar Geometry Probabilistic Interpretation

The points in each image are often subjected to a certain degree of noise or uncertainty, and even the inlier correspondences are not entirely accurate. Thus, the estimation of the Fundamental matrix and the epipolar lines geometry also contain a certain amount of uncertainty, depending on the image points' precision and distribution. Furthermore, each examined point is also subjected to some noise. Therefore, a probabilistic view that takes into account the fundamental matrix and the points' uncertainty is of need.

There has been some study in academic literature regarding the probability distributions of the Fundamental matrix and the epipolar constraint. Csurka et al. [24] first introduced the probabilistic formulation of the epipolar constraint and its properties. They derived the covariance matrix of the fundamental matrix, Σ_f, which represents the estimation uncertainty due to the uncertainty of the corresponding image points used for its estimation, and showed how to

compute it by both statistical (Monte-Carlo simulation) and analytical approaches.

Csurka et al. defined the *epipolar band* - the uncertainty region of the epipolar line. This is a region in which the probability of finding the match in one image of a given point in the other, is less than a certain value. The epipolar band of some probability p lies outside the conic $ll^T - k^2\Sigma_l$, where l is the epipolar line, Σ_l is its covariance matrix estimated from Σ_f using the covariance-propagation theorem, and k^2 is the square Mahalanobis distance corresponding to p. It can be shown that k^2 follows a χ^2 distribution with 2 degrees of freedom. An example can be found in Fig. 2.2, where the point marked by a red circle in Fig. 2.2a corresponds to some point within one of the limited regions (or bands) in Fig. 2.2b according to the predefined probabilities in the image legend (the epipolar bands in this figure were evaluated according to the points assignment in Fig. 1.1, with an assumed standard deviation of 5 pixels for each point).



(a) (b)

Figure 2.2: Epipolar band example. (a) The first image with a marked reference point. (b) Second image with some epipolar bands corresponding to different probabilities.

Sur et al. [25] utilized the same formulation for covariance estimation and computation of the epipolar band for the eight-point algorithm and demonstrated an application for robust estimation and outlier removal.

The epipolar band formulation can be used not only for the restriction of corresponding features to a given threshold, but also to validate and score given matches. Brandt [26] came up with an expression for the probability density function (PDF) of the point y in the second image given the point x in the first image, $P\left(y|x, \Sigma_f\right)$. His expression, allegedly, not only restricts the

matching point to the vicinity of the corresponding epipolar line, but also tells which area across the line is the most probable for it. However, the fact remains that the true location of a match along the epipolar line does not depend on the epipolar geometry nor the corresponding set known thus far, but rather depends on the scene itself which can be non smooth or discontinuous. What the suggested PDF really describes is the probability density, or the certainty, for the true epipolar line to pass through each image point. Brandt interpretation to the certainty of the corresponding point is the summation of the probabilities of all the lines passing through the point (lines that can explain it). By abuse of notation, the measure calculated by Brandt is:

$$\int P\left(l'|x, \Sigma_f, y \in l'\right) dl',$$

Which is the answer to the question: what is the marginalized probability for the data gathered thus far, given the new corresponding pair? The certainty of the point itself could not be inferred, though, in simple cases of smooth or near smooth surfaces, the true correspondence often lay in the proximity of the narrowest point of the epipolar band, as the PDF predicts. This occurs because the PDF is a manifestation of the known correspondences, laying on the same surface as the new point (in the example shown in Fig 2.2 above, this is not the case, as the true corresponding point does not lay in the narrowest region of the epipolar band, because of the discontinuous transformation between the images).

Another probabilistic measure for the quality of a match $\{x, y\}$ is suggested by Unger et al. [27]. Their measure is the probability associated with the Mahalanobis distance from the closest epipolar line explained by the point, to the mean epipolar line (the one estimated from the correspondences):

$$P\left(y \in l'|x, \Sigma_f\right) = F_{\chi_2^2}\left(k^2\right) \ , \tag{2.3}$$

where $F_{\chi_2^2}(\cdot)$ is the Chi-Square cumulative probability with 2 degrees of freedom, and

$$\begin{aligned} k^2 &= \min_m \ \left(l' - m'\right)^T \Sigma_{l'} \left(l' - m'\right), \\ s.t. \quad & m'\tilde{y} = 0. \end{aligned} \tag{2.4}$$

In the above, l' is the epipolar line associated with x, $l' = \tilde{x}\mathbf{F}$, and m' is the closest line to l', containing y. The k^2 value obtained by Eq. 2.4 yields $\tilde{y}^T\left(l'l'^T - k^2\Sigma_{l'}\right)\tilde{y} = 0$.

The probability in Eq. 2.3 can be interpreted as the probability of the corresponding pair being compliant with the epipolar geometry, $P\left(\{x, y\}\,|\mathbf{F}, \Sigma_f\right)$. The resulting measure is not a

probability distribution function for the match. It is not normalized for the entire image space, and does not tell the probability of y being the true correspondence of x. Nevertheless, it does give a proper score for the validity of the point match.

Also worth mentioning, in this context, is the work of Triggs [23] of Joint Feature Distribution (JFD). The formalism defined by Triggs is a general framework for probability distribution of corresponding features, in an affine or projective (homography or epipolar constrained) regime. As opposed to the aforementioned measures, this one does not incorporate the covariance of the fundamental matrix per se as described earlier, but is rather based on an empirical assumption about the probabilistic nature of the transformation given the corresponding points' scatter matrix. The proposed scheme yields an elliptical region of certainty (a Gaussian PDF) for each y given x. The densities are quite localized similarly to Brandt's PDF, and like it, JFD is also based on a heuristic assumption of continues and smooth enough surfaces for the captured scene. Of course this assumption may prove wrong in reality, where scenes are, in many cases, composed of several objects and background.

2.3 Graph Matching

In recent years, one of the deeply studied approaches for addressing the assignment problem is by means of graph matching (GM). One of the advantages of GM over robust model estimation is the fact that all candidates for each point can be tested, and not just the nearest ones. For practical reasons, the number of potential assignments is usually restricted to only a few candidate assignments for each point using the descriptors' metric. This may reduce both computational and storage burden, as well as restrict the search to a more feasible sub-space.

As opposed to the robust estimation techniques described above, graph matching technique is more flexible as it is not trying to enforce a given parametric model on the data, but rather to find a consistent set of points in one image which geometrically resembles the corresponding set in the other image. More precisely, it is designed to find an alignment which preserves the relations between groups of points (usually small groups of 2 to 3 points).

2.3.1 Pairwise Graph Matching

Given a set of points $S_1 = \{x_i\}_{i=1}^{N_1}$ in one image, and a set of points $S_2 = \{y_i\}_{i=1}^{N_2}$ in the second image, the Graph Matching aims to recover the proper assignment $C = \{c_{ii'}\}_{i=1}^{n}$, $n \leq \min\{N_1, N_2\}$,represented by an $N_1 \times N_2$ binary matrix, $\mathbf{Z}$, which entries are $z_{ij} = 1$ if x_i corresponds to y_j, and $z_{ij} = 0$ otherwise. This matrix, denoted as the *assignment matrix*, is row and column semi-normalized, meaning

$$\sum_{i=1}^{N_1} z_{ij} \leq 1 \ , \ \sum_{j=1}^{N_2} z_{ij} \leq 1.$$

In pairwise GM we intend to maximize the affinity between each pair of correspondences, given e.g. by the RBF (radial basis function) kernel,

$$\phi_2\left(c_{ii'}, c_{jj'}\right) \triangleq \phi_2\left(\{x_i, y_{i'}\}, \{x_j, y_{j'}\}\right) = \exp\left\{-\frac{1}{\varepsilon}\left|\|x_i - x_j\|_{L_2} - \|y_i - y_j\|_{L_2}\right|\right\}. \quad (2.5)$$

The value $\phi_2\left(c_{ii'}, c_{jj'}\right)$ is maximized when the distance between $\mathbf{x}_i$ and $\mathbf{x}_j$ is equal to the distance between $\mathbf{y}_{i'}$ and $\mathbf{y}_{j'}$, and is thus, an isometry affinity measure (see Fig. 2.3a). The parameter $\varepsilon > 0$ in Eq. 2.5 quantifies the tolerance to inexact isometry. This formulation can be thought of as a graph, which nodes are the correspondences $c_{ii'}$ and the edges connecting them are the affinities $\phi_2\left(c_{ii'}, c_{jj'}\right)$.

In classical GM approaches, the graph is assumed to consist of a strongly connected cluster, representing the true correspondences - each yielding high affinities, and some outliers, that yield high affinities only at random. The correspondence problem is equivalent to finding the cluster C maximizing the sum of corresponding pairwise affinities.

$$C^* = \arg\max_{C} \sum_{c_{ii'}, c_{jj'} \in C} \phi_2\left(c_{ii'}, c_{jj'}\right). \quad (2.6)$$

It is convenient to define a matrix representation of the graph in the form of the *affinity matrix* $\mathbf{A} \in \mathbb{R}^{N_1 \cdot N_2 \times N_1 \cdot N_2}$

$$\mathbf{A}\left(i, i'; j, j'\right) \triangleq \mathbf{A}\left(i\left(N_2 - 1\right) + i', j\left(N_2 - 1\right) + j'\right) = \phi_2\left(c_{ii'}, c_{jj'}\right). \quad (2.7)$$

The optimal assignment is thus given by

$$\begin{aligned} \mathbf{z}^* &= \arg\max_{\mathbf{z}} \left(\mathbf{z}^T \mathbf{A} \mathbf{z}\right) \\ s.t. \quad & \mathbf{z} \in \{0,1\}^{N_1 N_2}, \\ & \sum_i z_{ij} \leq 1 \qquad \forall j, \\ & \sum_j z_{ij} \leq 1 \qquad \forall i \end{aligned} \quad (2.8)$$

where $\mathbf{z}$ is a row-stacked replica of the assignment matrix $\mathbf{Z}$, which entries are z_{ij}.

This optimization problem is a *quadratic binary problem* (QBP), which unfortunately, known to be NP-hard. Therefore, it is commonly approximated by computationally efficient schemes.

The Graduated Assignment (GA) algorithm by Gold and Rangarajan [14] relaxes the constraint $\mathbf{z} \in \{0,1\}^{N_1 N_2}$ in Eq. 2.8 to $\mathbf{w} \in \mathbb{R}^{N_1 N_2}$. It applies an iterative annealing scheme for the maximization of the aforementioned objective, combined with an entropy penalty term, updating the derivative of the expression at each iteration.

Leordeanu et al. [15] proposed a spectral relaxation of Eq. 2.8,

$$\begin{aligned} \mathbf{w}^* &= \arg\max_{\mathbf{w}} \left(\frac{\mathbf{w}^T \mathbf{A} \mathbf{w}}{\mathbf{w}^T \mathbf{w}} \right), \\ s.t. \qquad & \mathbf{w} \in \mathbb{R}^{N_1 N_2}, \end{aligned} \tag{2.9}$$

for their Spectral Matching (SM) algorithm. Equation 2.9 is solved by computing the principal eigenvector of $\mathbf{A}$, and applying a discretization scheme to $\mathbf{w}^*$ to compute an approximate solution of Eq. 2.8.

The same authors later suggested an improvement to the discretization step [16], which can also be applied by its own as a full matching algorithm. Their Integer Projected Fixed Point (IPFP) algorithm iteratively projects the proposed solution into the discrete space defined by the constraints of Eq. 2.8. IPFP is claimed to outperform previous approaches.

Cho et al. [17] proposed Reweighted Random Walk Matching (RRWM), a different approach inspired by algorithms from the field of Web-ranking. They define an "affinity-preserving random walk" on the graph, by replacing the need for stochastic row-wise normalization with an absorbing node, yielding a formulation identical to the spectral relaxation of Leordeanu et al. [15]. The crux of their approach is the adaptation of the personalization approach from Web-ranking algorithms, dynamically reweighing the edges of the graph, in order to induce the matching constraints in the course of the random walk itself.

2.3.2 Higher Order Graph Matching

More advanced methods use higher order affinities (usually third order) to allow for more general transformations. It has been shown [18–20] that even when a small change in scale is induced between the two images, the pairwise GM algorithms' performances are considerably

degraded. In order to deal with more complex transformations, triplets matching is considered. In triplets GM, a hypergraph is constructed, in which each node represents a correspondence $c_{ii'}$, and each hyper-edge connecting three of them represents an affinity measure preserved by scale change. This affinity can be, for example,

$$\phi_3\left(c_{ii'}, c_{jj'}, c_{kk'}\right) \triangleq \phi_3\left(\{x_i, y_{i'}\}, \{x_j, y_{j'}\}, \{x_k, y_{k'}\}\right) = \exp\left\{-\frac{1}{\varepsilon}\left|\|\theta_{ijk} - \theta_{i'j'k'}\| + \|\theta_{jki} - \theta_{j'k'i'}\| + \|\theta_{ikj} - \theta_{i'k'j'}\|\right|\right\}, \quad (2.10)$$

where θ_{ijk} is the angle formed by the three points x_i, x_j and x_k. The triplets affinity measure $\phi_3\left(c_{ii'}, c_{jj'}, c_{kk'}\right)$ is maximized when the difference between corresponding angles is minimized, i.e. the two triangles, defined by x_i, x_j, x_k and $y_{i'}, y_{j'}, y_{k'}$ respectively, are approximately similar. Therefore, the triplets affinity measure defined above is a scale-invariant feature (see Fig. 2.3b).

The expansion for higher order affinities is non-trivial. Several recent papers suggest different techniques for addressing the issue. Zass et al. [18] derive a probabilistic formulation for hypergraph matching (which will be denoted as HGM throughout this paper), showing that the calculated hypergraph matrix can be defined probabilistically as a successive Kronicker product of the one-dimensional probability vector. This observation results in a convex optimization scheme for recovering this vector, which constitutes a soft assignment matching.

Duchenne et al. [20] use similar formulation. In their work the high order affinity hypergraph is represented by a tensor, and the solution to the optimization problem is derived by extending the spectral relaxation of Leordinue et al. [15] to higher dimensions. This is done by generalizing the power iteration matrix eigendecomposition algorithm for the case of multi-way tensors (this algorithm will be denoted as TGM - Tensor Graph Matching).

Another work dealing with high order matching was conducted by Chertok et al. [19]. This work (denoted in this thesis as THOA - Tensor High Order Assignment) is pretty close to the former two, with one significant difference: The high order affinity tensor is not stored in memory nor processed as a whole. Instead, the tensor is marginalized to a second order matrix, solved using spectral matching [15]. This paradigm results in a much efficient solution. Further improvement of time and storage complexity is achieved by sparsifying the tensor, based on results from the field of random matrix theory.

2.3.3 Probabilistic Graph Matching

Another important aspect of GM methods is their probabilistic interpretation if exists. Some of the described works [17, 18] as well as the work done by Egozi et al. [28] adopt a probabilistic interpretation for the graph matching problem, inducing a soft probability assignment vector as the output of the algorithm. The probabilistic formulation is derived by assuming the following working assumptions: (1) The pairwise (or triplets) affinities can be seen as an empirical estimate for the pairwise (or triplets resp.) assignment probabilities: $\phi\left(c_{ii'}, c_{jj'}, ...\right) \approx P\left(c_{ii'}, c_{jj'}, ...\right)$. (2) The pairwise assignments of different points are statistically independent:

$$P\left(c_{ii'}, c_{jj'}\right) = P\left(c_{ii'}\right) P\left(c_{jj'}\right). \tag{2.11}$$

Given those heuristic justified assumptions, some of the proposed GM methods can be interpreted probabilistically. The spectral relaxation [15], for example, can be seen as an approximation for the assignment probabilities, as the spectral relaxation of $\mathbf{A}$ is its Rank-One-Approximation (ROA) in terms of the Frobenius norm:

$$\mathbf{p} = \mathbf{w}^* = \arg\max_{\mathbf{w}} \left(\frac{\mathbf{w}^T\mathbf{A}\mathbf{w}}{\mathbf{w}^T\mathbf{w}}\right) = \arg\min_{\mathbf{w}} \left\|\mathbf{A} - \mathbf{w}\mathbf{w}^T\right\|_{L_2}. \tag{2.12}$$

As $\mathbf{A}$ is symmetric and nonnegative, by Perron-Frobenius theorem, we have that $\mathbf{p}$ is nonnegative as well, as expected from a probabilistic measure.

Egozi et al. [28] extended the probabilistic graph matching formulation to utilize unary probabilities $P\left(u\left(c_{ii'}\right)\right)$, given by the marginal assignment probabilities estimated in Eq. 2.12. They derived an iterative probabilistic scheme, where at each iteration the estimated unary probabilities are used to boost the matching process, by improving both estimations of the unary probabilities, $P\left(u\left(c_{ii'}\right)\right)$, and the conditional assignment probabilities matrix, $P\left(c_{ii'}|c_{jj'}\right)$.

In this work it is shown that other sources of information, such as geometrical priors, or the similarity of corresponding interest points, modeled by logistic regression, can be used to quantify unary probabilities $P\left(u\left(c_{ii'}\right)\right)$, and by incorporating them into the graph, improve the GM performance. We demonstrate this concept by estimating the unary probabilities via the epipolar geometry.

2.3.4 Multiple Image Registrations via Multiple Eigenvectors

Chertok et al. [29] presented a spectral graph scheme for symmetry analysis (SSA) that utilizes multiple eigenvectors. Different symmetry axes are recovered by matching an image to itself, and estimating the resulting multiple image-to-image assignments, each corresponding to a particular symmetry axis. Thus, each such assignment is a maxima of the Reighley quotient in Eq. 2.9, and can be recovered by computing multiple leading eigenvectors $\left\{\mathbf{w}^{(m)}\right\}_{m=1}^{M}$, instead of just one. This is of particular interest in the context of the proposed work, as we claim that the geometrical structure of wide-baseline stereo setup often implies that they might include several distinct motions due to the varying depth of the objects with respect to the cameras' planes.

The downside of Chertok's approach is that nonnegativity is only guaranteed for the leading eigenvector by the Perron-Frobenius theorem. The orthogonality of the eigenvectors implies that the succeeding eigenvectors will either have negative entries, or disjoint support, implying that they encode the matching of spatially disjoint regions in the image (which is hardly ever the case).

2.4 Discussion

In this chapter we reviewed the two main approaches for dealing with the assignment problem, the epipolar geometry robust estimation approach, and the graph matching approach. Both have strengthes and weaknesses, and in both the performances degrade as the input scenario approaches the wide-baseline domain, with non-negligible perspective deformations.

The epipolar robust estimation techniques are only designed to work when the input images were captured by a pinhole-model cameras. Their main strength is in being rigorous and general, making them suitable for every stereo pair regardless of the perceived scene. However, their generality is also a major drawback. Since epipolar robust estimation does not enforce any rigid transformation, the points in one image can fit, according to the constraint 2.1, any other point within the epipolar line. This could result in a massive clutter, especially when the captured scene contains repetitive patterns in a direction parallel to the baseline (which is often the case when shooting buildings for example). These techniques do not use any smoothness or piecewise rigidness assumptions, which could benefit the matching process by mitigating this

behavior. Another drawback of robust estimation approach is its inability to cope with more than just one nearest neighbor.

Graph matching algorithms, on the other hand, are not so rigorous regarding the model constraint, and so, are often used for matching non-rigid objects such as faces. They heuristically assume a global relaxed but rigid motion model, and thus cope well with narrow-baseline scenarios. Nevertheless, in wide-baseline cases this assumption does not hold, as the deformations can disjoint parts of the scene causing for non rigidity of the transformation. In addition, the GM approach is sensitive to other erroneous alignments, unrelated to camera transformation, such as symmetries or repetitive objects in the scene. Such alignments could clutter the results of the schemes.

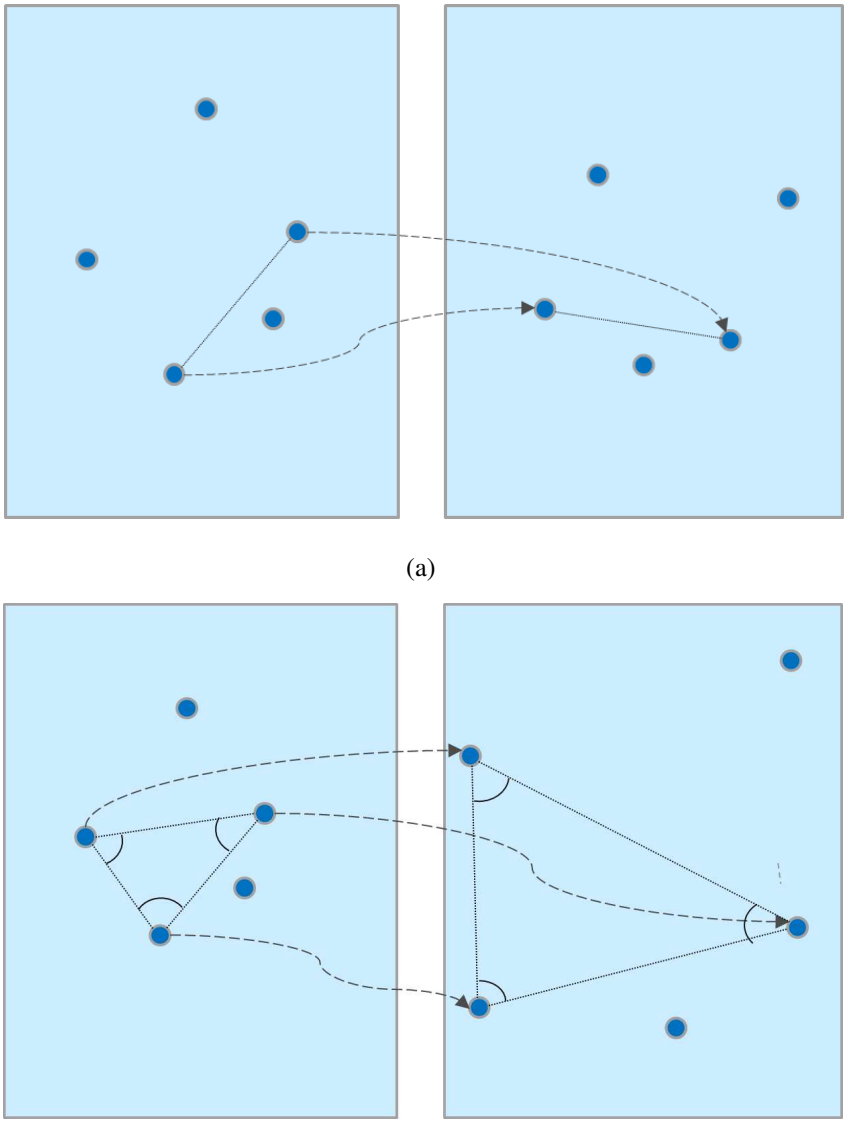


Figure 2.3: Matching affinities of different orders. (a) Pairwise matching affinity - The maximization of the score induces an isometry transform - Equal distance between corresponding pairs. (b) Triplets matching affinity - The maximization of the score induces a similarity transform. Each triangle formed from three points in one image is similar to the corresponding triangle in the second image.

Chapter 3

A Unified Wide-Baseline Stereo Matching Approach

3.1 Introduction

Although GM and robust estimation approaches suggest powerful tools for dealing with the assignment problem in the presence of noise and outliers, both methodologies lack when facing a wide-baseline image matching scenario. The registration of wide-baseline stereo images is often difficult, as both local and global perspective deformations degrade the inlier rate and the global motion model simplicity. In such cases, the epipolar robust estimation techniques often prove insufficient due to lack of inliers in the initial putative set of correspondences, and repetitive patterns along the baseline direction. On the other hand, the GM assumption of one global approximately rigid transformation, is often invalid for a wide-baseline setup. Thus in this case, GM performances are often considerably damaged too.

The two images of the wide-baseline stereo pair might be related by multiple spatial transformations, each related to image components with different distance form the camera plane. This is exemplified in Fig. 1.1 where the Sphinx and pyramid appear in both images and are related by different spatial transformations.

This observation motivates our multi-assignment GM approach. The proposed algorithm preserves GM strength of robust and efficient cluster matching, by first generalizing its model to match several components instead of just one. It then combines all components probabilistically, and by introducing the epipolar probability into the affinity graph, enforces the global

constraint, and attenuate false correspondences.

This chapter is organized as follows: In section 3.2 we present the multi-assignment framework, and derive the optimization scheme for calculating the subgraph components. Section 3.3 introduces the combined probabilistic approach, which binds together the epipolar constraint with the graph affinity probabilities. The Multi-Assignment Graph Matching Algorithm (MAGMA) is described in detail in Sec. 3.4. The chapter is concluded in Sec. 3.5.

3.2 Multi-Assignment Graph Matching for Image Registration

In contrast to common graph matching approaches we aim to recover multiple assignments, rather than a single one. The core of our approach is that each relative motion will be manifested by a weakly connected subgraph (see Fig. 3.1) within the graph matching framework. Given two images, I_1 and I_2, and the sets $S_1 = \{x_i\}_{i=1}^{N_1}$ and $S_2 = \{y_i\}_{i=1}^{N_2}$, such that $x_i \in I_1 \; \forall i$,

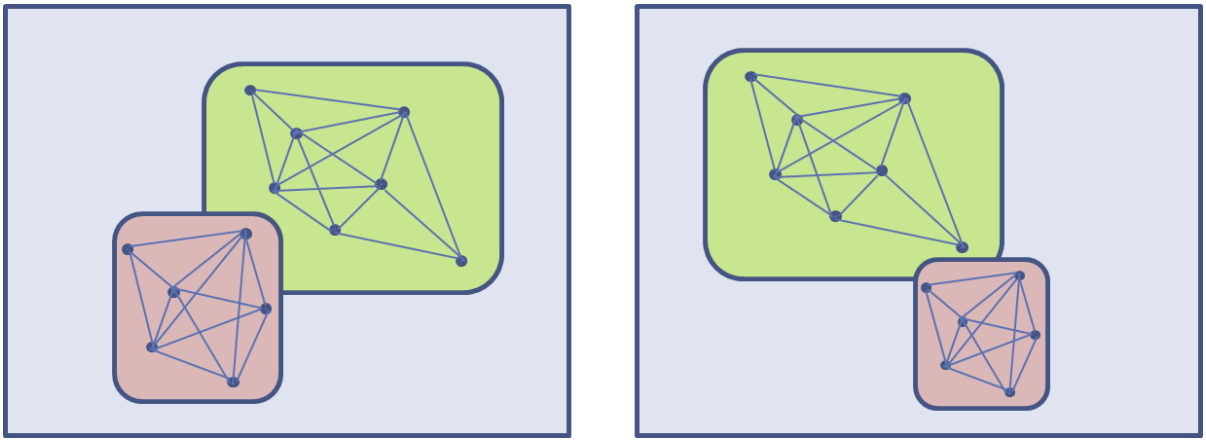


Figure 3.1: The multi-cluster model: Several objects in the scene (including e.g. the background and foreground of the scene). Each having a different transformation between the two images, and each can be represented as a cluster (or sub-graph).

and $y_i \in I_2 \; \forall i$, we construct our model similarly to the GM model. Our goal is to find the top M maximal consistent components in the graph, or the M sets of matches $\left\{C^{(1)}, C^{(2)}, ..., C^{(M)}\right\}$, which best explain the observed affinities, as depicted in the matrix $\mathbf{A}$. To formulate the problem, each set $C^{(m)}$ will be associated with a soft assignment matrix $\mathbf{W}^{(m)} \in \mathbb{R}^{N_1 \times N_2}$, which

entries stand for the pairwise matching probabilities in accordance to the appropriate component in the graph (as described for a single component by Egozi et al. [28]). The matrices $\mathbf{W}^{(m)}$ must satisfy the following constraints:

$$\begin{aligned} &\left[\mathbf{W}^{(m)}\right]_{ij} \in [0,1] \quad \forall m,\ i,\ j, \\ &\sum_i \left[\mathbf{W}^{(m)}\right]_{ij} \leq 1 \quad \forall m,j, \\ &\sum_j \left[\mathbf{W}^{(m)}\right]_{ij} \leq 1 \quad \forall m,i \end{aligned} \tag{3.1}$$

The inequalities are due to the fact that not all points in one image have true correspondences in the second image, and vice-versa.

Next we introduce the *soft multi-assignment matrix*, $\mathbf{\Omega} \in \mathbb{R}^{N\times M}$ ($N = N_1 \cdot N_2$), in which the columns $\left\{\mathbf{w}^{(m)}\right\}_{m=1}^{M}$ are the row-stacked replicas of $\left\{\mathbf{W}^{(m)}\right\}_{m=1}^{M}$: $\mathbf{\Omega} = \begin{pmatrix} \mathbf{w}^{(1)} & \mathbf{w}^{(2)} & \dots & \mathbf{w}^{(M)} \end{pmatrix}$.

we are aiming to solve the following decomposition problem:

$$\mathbf{A} \approx \mathbf{\Omega}\mathbf{\Omega}^T = \mathbf{w}^{(1)}\mathbf{w}^{(1)T} + \mathbf{w}^{(2)}\mathbf{w}^{(2)T} + \dots + \mathbf{w}^{(M)}\mathbf{w}^{(M)T}. \tag{3.2}$$

Equation 3.2 can be seen as a partitioning operation on the affinity graph, extracting its most prominent components. Each component contributes a part of the graph (one addend in Eq. 3.2), and the whole graph is completely composed by the sum of those contributions. This formulation is consistent with our model of several unconnected objects in the scene, each associated with a distinct sub-graph. The graph in total is mainly composed of these components, and some residual unstructured, weakly connected parts which do not agree with any of them, and would not affect the probabilities much.

This model of several unconnected components in the graph is not always valid, and even when it is, the actual number of components in the graph is unknown. Yet, a proper probabilistic derivation will allow for soft partitioning of the graph. Followed by simultaneous discretization of all components, it will assure each point correspondence will be taken according to the most appropriate assignment group.

The decomposition 3.2 can be rewritten as an optimization problem:

$$\begin{aligned} \mathbf{\Omega}^* = \arg\min_{\mathbf{\Omega}} & \left\|\mathbf{A} - \mathbf{\Omega}\mathbf{\Omega}^T\right\|_{L_2} \\ s.t. \quad & \mathbf{\Omega} \in [0,1]^{N\times M}, \\ & \sum_i \left[\mathbf{W}^{(m)}\right]_{ij} \leq 1 \qquad \forall m,j, \\ & \sum_j \left[\mathbf{W}^{(m)}\right]_{ij} \leq 1 \qquad \forall m,i \end{aligned} \tag{3.3}$$

The problem 3.3 is similar to the one being solved by the SSA algorithm (see Sec. 2.3.4). In order to extract M different components in the graph, the SSA factorizes the matrix $\mathbf{A}$ by using order-M spectral decomposition, known to solve the following decomposition problem:

$$\begin{aligned} \mathbf{\Omega}^* = \arg\min_{\mathbf{\Omega}} & \left\|\mathbf{A} - \mathbf{\Omega}\mathbf{\Omega}^T\right\|_{L_2} \\ s.t. \quad & \mathbf{\Omega}^T\mathbf{\Omega} = \mathbf{I} \end{aligned} \tag{3.4}$$

3.3 and 3.4 are both generalization of the classical SM problem [15], as both, for the case of $M = 1$, solve the Rank-One-Approximation (ROA) problem. Both formulations can be seen as best Rank-M-Approximations (RMA) under different constraints.

For $M = 1$, the Perron-Frobinius theorem guarantees that the leading eigenvector of $\mathbf{A}$ exists and is nonnegative, which allows the convenient discretization of the result (e.g. by greedy method), and provides a nice probabilistic interpretation to it. However, the same can not be said about other eigenvectors of $\mathbf{A}$. On the contrary, for $M > 1$, the orthogonality constraint, $\mathbf{\Omega}^T\mathbf{\Omega} = \mathbf{I}$, dictates that all other eigenvectors are orthogonal, and thus contain negative entries, making the results hard to interpret. For this reason, it can be inferred that this formulation is not the most appropriate one, neither for discretization nor for probabilistic interpretation.

Another known method for solving a similar problem is Symmetric Nonnegative Matrix Factorization (NMF) [30–32]. NMF is a well known established clustering technique used in recent years for several applications, including: face recognition [31], documents clustering [33, 34], and audio blind source separation [35].

Symmetric NMF solves the problem:

$$\begin{aligned} \mathbf{\Omega}^* = \arg\min_{\mathbf{\Omega}} & \left\|\mathbf{A} - \mathbf{\Omega}\mathbf{\Omega}^T\right\|_{L_2} \\ s.t. \quad & \mathbf{\Omega} \geq \mathbf{0} \end{aligned} \tag{3.5}$$

This problem can be solved efficiently either by a multiplicative iterative algorithm [36], or a nonnegative projected alternating least squares (ALS) approach [32].

Symmetric NMF allows for an intuitive discretization and a meaningful probabilistic interpretation, similarly to the one-component case (such as in SM), as it ensures nonnegativity of all assignment vectors' components.

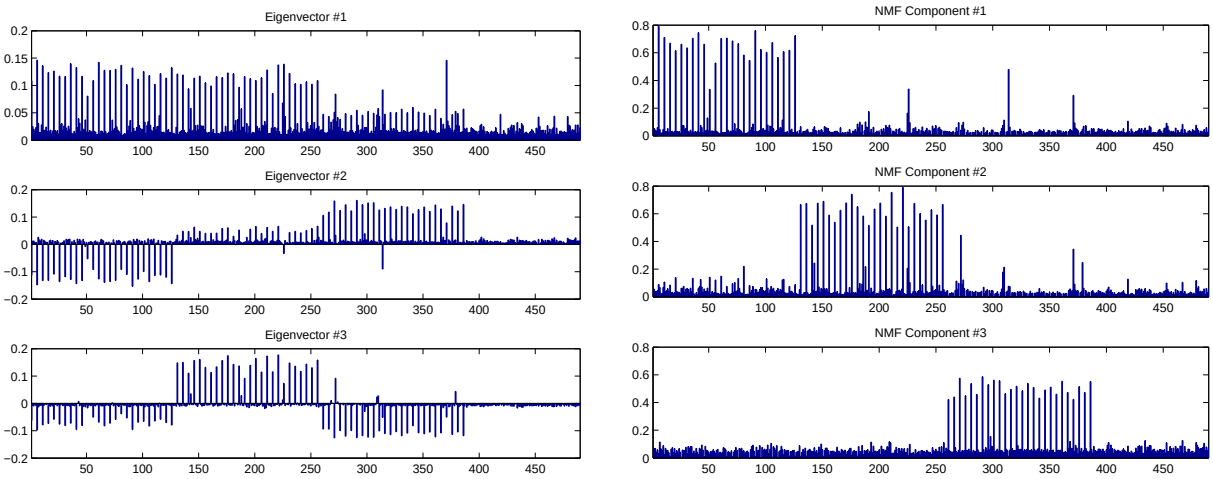
To show the advantages of Symmetric NMF over spectral decomposition, a simulation was conducted: A set of uniformly distributed points in 2D space was generated. From these points 3 groups were undergone different similarity (scale and rotation) transformations and perturbed with a random distortion. To each point additional candidate correspondences were generated at random. Some outlier points were given only randomly generated candidates. This setup initiated a graph using a scale invariant marginalized tensor [19]. Fig. 3.2 shows both spectral decomposition results, depicted by the three leading eigenvectors (a), and Symmetric NMF results of order 3 (b). It is quite noticeable that the spectral decomposition is not suitable for the job, as it yields hard-to-interpret vectors, which do not preserve the real nature of the graph. On the other hand, the Symmetric NMF results are much easier to interpret and present three distinguished components. The Symmetric NMF results can be easily combined and simultaneously discretized to form a hard assignment, using e.g. greedy approach. As all of the components originated from a single matrix optimization problem, no additional normalization between the components is needed.

Though the Symmetric NMF method yields much better results than the SSA approach, it still does not solve the problem stated by 3.3, as it does not enforce the (semi) stochastic constraints: $\sum_i \left[\mathbf{W}^{(m)}\right]_{ij} \leq 1$, $\sum_j \left[\mathbf{W}^{(m)}\right]_{ij} \leq 1$. We introduce here the Stochastic Constrained Matrix Factorization (SCMF) algorithm, a scheme for solving Eq. 3.3. Our scheme uses an approach similar to the projected ALS used for solving the Symmetric NMF problem, with a projection operator which orthogonally projects the result of each step to the feasible domain of the constraints 3.1. The nonnegative-projected-ALS method for calculating NMF was shown to be equivalent to projected-Newton optimization scheme [32], and the same is also true for SCMF. Next we derive the SCMF projection step.

The constraints define the following linear subspace:

$$Q = \{\mathbf{\Omega} : \mathbf{\Psi}\mathbf{\Omega} - \mathbf{\Lambda} \leq \mathbf{0}\},$$

where $\mathbf{\Psi}$ and $\mathbf{\Lambda}$ define the linear constraints: nonnegativity, rows (semi) stochastic normaliza-



(a) (b)

Figure 3.2: Multi assignment graph matching - several solutions from one affinity graph. (a) Spectral decomposition. The three leading eigenvectors are displayed. Each eigenvector is associated with one component in the graph. (b) Symmetric NMF. Each column in the factorization is associated with different component.

tion and columns (semi) stochastic normalization.

$$\mathbf{\Psi} = \begin{pmatrix} -\mathbf{I}_{N\times N} \\ \mathbf{\Psi}^{\text{rows}}_{N\times N} \\ \mathbf{\Psi}^{\text{cols}}_{N\times N} \end{pmatrix}, \mathbf{\Lambda} = \begin{pmatrix} \mathbf{0}_{N\times M} \\ \mathbf{1}_{N\times M} \\ \mathbf{1}_{N\times M} \end{pmatrix} .$$

The matrices $\mathbf{\Psi}^{\text{rows}}_{N\times N}$ and $\mathbf{\Psi}^{\text{cols}}_{N\times N}$ are sparse matrices, encoding the rows and columns constraints respectively. Thus the orthogonal projection operator to the subspace Q is given in the form:

$$P_Q\left(\mathbf{\Omega}\right) = \mathbf{\Omega} - \mathbf{\Psi}^T\left(\mathbf{\Psi}\mathbf{\Psi}^T\right)^{-1}\left[\mathbf{\Psi}\mathbf{\Omega} - \mathbf{\Lambda}\right]_+ , \tag{3.6}$$

where the operator, $[\ \cdot\ ]_+$, is the nonnegative projection operator, defined as: $[\mathbf{\Omega}]_+ = \max\{\mathbf{\Omega}, 0\}$. The SCMF scheme is summarized in Alg. 1. An additional scaling factor λ is introduced to regularize the iterative algorithm convergence (we used $\lambda = 0.5$).

3.3 Unified Probabilistic Stereo Image Matching

In this work we propose a unified approach for wide-baseline image matching, that utilizes both graph matching (Section 2.3) and RANSAC (Section 2.2.2) via a joint probabilistic formulation.

Algorithm 1 Symmetric Semi-Stochastic Constrained Matrix Factorization (SCMF)

Input: A symmetric nonnegative matrix, $\mathbf{A} \in {\mathbb{R}_+}^{N \times N}$,

The factorization order $M \ll N$

Output: Solution to the optimization problem 3.3, $\mathbf{\Omega} \in {\mathbb{R}_+}^{N \times M}$

1: Calculate the constraint coefficients: $\mathbf{\Psi}$ and $\mathbf{\Lambda}$

2: Set the regularization scaling parameter λ (e.g. $\lambda = 0.5$)

3: Initialize $\mathbf{\Omega}$ randomly

4: $\mathbf{\Omega} \leftarrow P_Q(\mathbf{\Omega}) = \mathbf{\Omega} - \mathbf{\Psi}^T \left(\mathbf{\Psi}\mathbf{\Psi}^T\right)^{-1} [\mathbf{\Psi}\mathbf{\Omega} - \mathbf{\Lambda}]_+$

5: **repeat**

6: $\quad \mathbf{\Omega}_0 = \mathbf{\Omega}$

7: $\quad \mathbf{\Omega} \leftarrow \mathbf{\Omega}\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}$

8: $\quad \mathbf{\Omega} \leftarrow P_Q(\mathbf{\Omega}) = \mathbf{\Omega} - \mathbf{\Psi}^T \left(\mathbf{\Psi}\mathbf{\Psi}^T\right)^{-1} [\mathbf{\Psi}\mathbf{\Omega} - \mathbf{\Lambda}]_+$

9: $\quad \mathbf{\Omega} \leftarrow (1-\lambda)\mathbf{\Omega}_0 + \lambda\mathbf{\Omega}$

10: **until** convergence

Let I_1 and I_2 be the input images, and $S_1 = \{x_i\}_{i=1}^{N_1}$ and $S_2 = \{y_i\}_{i=1}^{N_2}$ be the corresponding sets of feature points.

We formulate the matching as a multivariate estimation problem, whose set of variables is divided into two sets of parameters. The first of which are the assignment probabilities $\mathbf{p}$ computed in Eq. 2.12, which encode the geometrical similarity between the sets of feature points in both input images, and are estimated by graph matching. The second set of parameters encodes the epipolar geometry in terms of the fundamental matrix $\mathbf{F}$, and is estimated via the RANSAC scheme or one of its derivations as in Section 2.2.

The gist of our approach is that using an estimate of the point matching given by the assignment probabilities $\mathbf{p}$, one can use $\mathbf{p}$ to find the most probable set of point matches between I_1 and I_2. This matching allows to efficiently apply the RANSAC approach to estimate the epipolar geometry parameters $\mathbf{F}$, relating I_1 and I_2. Furthermore, using $\mathbf{F}$ one can estimate the unary assignment probabilities, by utilizing the epipolar probability compliance measure [27], given by Eq. 2.3, and improve the graph matching accuracy as will be shown next.

In order to utilize the unary assignment probabilities within the graph matching scheme, we reformulate Eq. 2.5 (the same derivation can also be applied on marginalized probabilities

of any order, e.g. by marginalizing Eq. 2.10). Following Bayes' theorem, and the unary probabilities independence assumption:

$$\begin{aligned} P(c_{ii'}, c_{jj'}) &= P(c_{ii'}, c_{jj'} | u(c_{ii'}), u(c_{jj'})) P(u(c_{ii'}), u(c_{jj'})) \\ &= P(c_{ii'}, c_{jj'} | u(c_{ii'}), u(c_{jj'})) P(u(c_{ii'})) P(u(c_{jj'})), \end{aligned} \tag{3.7}$$

where,

$$P(c_{ii'}, c_{jj'} | u(c_{ii'}), u(c_{jj'})) = \exp\left\{ -\frac{1}{\varepsilon} \left| \|x_i - x_j\|_{L_2} - \|y_i - y_j\|_{L_2} \right| \right\}. \tag{3.8}$$

Equation 2.5 quantifies the pairwise probability solely by the geometrical similarity of edges, overlooking the cases where edges of similar length are non-corresponding. In contrast, Eqs. 3.7 and 3.8 provide finer statistical modeling of point matching. By setting $P(u(c_{ii'})) = 1$ $\forall i$, Eq. 3.7 is reduced to Eq. 2.5. Alternatively, a supplementary source of information could contribute to the refinement of the probabilistic model. We estimate the independent unary probabilities by the epipolar compliance probability an in Eq. 2.3.

This yields an iterative registration scheme that alternates between running graph matching and robust epipolar geometry estimation (e.g. by RANSAC), and we formulate the proposed scheme via the likelihood maximization

$$(\mathbf{p}^*, \mathbf{F}^*) = \arg\max_{\mathbf{p},\mathbf{F}} L(\mathbf{p}, \mathbf{F}) = \arg\max_{\mathbf{p},\mathbf{F}} L_G(\mathbf{p}, \mathbf{F}) + L_E(\mathbf{p}, \mathbf{F}) \tag{3.9}$$

where $L_G(\mathbf{p}, \mathbf{F})$ is the likelihood of the probabilistic graph matching, and $L_E(\mathbf{p}, \mathbf{F})$ is the likelihood of the epipolar model. It was shown by Egozi et al. [28] that the Spectral Graph Matching scheme computes the maximum likelihood estimate of the matching problem and thus maximizes $L_G(\mathbf{p}, \mathbf{F})$. Similarly, the robust estimation process evaluates the maximum likelihood estimate of $\mathbf{F}$ from inlier assignments, and thus maximizing $L_E(\mathbf{p}, \mathbf{F})$. This paves the way for the derivation of an iterative, non-decreasing scheme that iterates

$$\mathbf{p}_n = \arg\max_{\mathbf{p}} L_G(\mathbf{p}_{n-1}, \mathbf{F}_{n-1}) \tag{3.10}$$

$$\mathbf{F}_n = \arg\max_{\mathbf{F}} L_E(\mathbf{p}_n, \mathbf{F}_{n-1}) \tag{3.11}$$

and it follows that

$$L_G(\mathbf{p}_n, \mathbf{F}_{n-1}) \geq L_G(\mathbf{p}_{n-1}, \mathbf{F}_{n-1}) \tag{3.12}$$

and

$$L_E(\mathbf{p}_n, \mathbf{F}_n) \geq L_E(\mathbf{p}_n, \mathbf{F}_{n-1}) \tag{3.13}$$

Thus,

$$L(\mathbf{p}_n, \mathbf{F}_n) \geq L(\mathbf{p}_n, \mathbf{F}_{n-1}). \tag{3.14}$$

The proposed framework can be generalized, to use other probabilistic sources of information, such as appearance similarity (e.g. the use of SIFT based similarity measures in PROSAC [11], BEEM [12] or BLOGS [13]), other orders of graph alignments or some knowledge of parametric transformation between the two images (e.g. JFD [23]).

3.4 Implementation

Now, using the principles described in former sections, we can derive our full algorithm, the *Multi-Assignment Graph Matching Algorithm* (MAGMA). Its details are as follows: Since epipolar geometry is unknown to begin with, our approach consists of iterative scheme in which the epipolar model inferred from each iteration is used to enhance the next one, like described in the previous section.

We start by using the original pairwise (or marginalized higher order) measures to construct the affinity matrix $\mathbf{A}$ (as depicted in Eq. 2.7). Next SCMF decomposition scheme is used for gathering the M components of assignments, and all of them are discretized simultaneously, enforcing the original constraints of Eq. 2.8.

In our experiments we used marginalized tensor calculated from angles (third-order matching, see Eq. 2.10) in order to get each assignment component to approximate a similarity (i.e. scale invariant) transformation. The marginalized tensor was decomposed using SCMF (Alg. 1), where the factorization order (i.e. number of components) was set to be $M = 3$. This number generally gives good results. Other values for M have also been tested, and gave similar results. Methods for inferring the real number of components are also addressed in literature, however they are mostly parametric and not very consistent. That is why we have used a constant number of components, which in our opinion should give good results, as it is not too big (which may cause overfitting or over-partitioning of assignments), and on the other hand sufficient for most natural scenes and images.

After the initial global assignment was obtained, the epipolar geometry is inferred using robust estimation technique. We have used RANSAC with local optimization step (of constant number of iterations) for each new optimum obtained. In order to accelerate the process, if the points were originated from scale invariant features such as SIFT, the scale and orientation information is used to obtain each global RANSAC iteration from only two samples, in a fashion similar to the BEEM algorithm [12]. Local steps (and global ones in case no scale and orientation data was available) are calculated using the normalized-eight-point algorithm [21]. The points' STD (standard deviation), σ, is then being calculated using the Sampson distance [6] between the inlier points and their respective epipolar lines, and in turn σ is used for the evaluation of the fundamental matrix covariance, Σ_f [24, 25].

After acquiring the fundamental matrix and its uncertainty, the affinity matrix is updated accordingly, utilizing the correspondences' unary epipolar probabilities as in Eq. 3.7.

This process continues iteratively for a $MaxIter$ number of iterations, or until a stopping criteria is met, when the inliers' STD, σ, reaches a predefined threshold, TH_σ, or the relative change gets bellow a certain value, TH_{ratio}.

The whole process usually takes two to three iterations to converge. Since the calculation of the third order affinities is the most time-consuming part of the algorithm (and it is only done once), the algorithm's time complexity is about the same as other tensor based matching schemes. The scheme is generally outlined in Alg. 2.

3.5 Summary

We have presented a complete framework for matching stereo images, which is especially designed for the wide-baseline case. The proposed algorithm includes several steps for ensuring its applicability to the wide-baseline scenario, and robustness to outliers. The algorithm generalizes GM approach to better cope with non rigid deformations between the input images, by allowing for more than one assignment to be calculated from the affinity graph. In addition, it considers epipolar geometry in a probabilistic way, to further improve accuracy and robustness.

Algorithm 2 Multi-Assignment Graph Matching Algorithm (MAGMA)

Input: Image points from both images, $S_1 = \{x_i \in I_1\}_{i=1}^{N_1}$ and $S_2 = \{y_i \in I_1\}_{i=1}^{N_2}$, $N_1 \times k_{nn}$ matrix, mapping each point in S_1 to its k_{nn} nearest neighbors in S_2

Output: Correspondence map $C = \{c_{ii'}\}_{i=1}^{n}$ between S_1 and S_2

Construct the affinity matrix $\mathbf{A}$ from the given S_1, S_2 and the nearest neighbor table (according to Eq. 2.10)
while (number of repetitions $\leq MaxIter$) **do**
  Factorize $\mathbf{A}$ Using SCMF (Algorithm 1): $\mathbf{A} \approx \mathbf{\Omega}\mathbf{\Omega}^T$
  Discretize $\mathbf{\Omega}$ to get putative set of matches
  Robustly estimate the epipolar geometry to get $\mathbf{F}$ and a set of inlier matches
  Use inliers to estimate the points' STD, σ, and fundamental matrix covariance, Σ_f
  if $(\sigma < TH_\sigma)$ **or**$(\sigma(prev)/\sigma(curr) < TH_{ratio})$ **then**
    return: Current inliers
  end if
  Calculate epipolar compliance for each putative match: $P\left(u\left(c_{ii'}\right)\right) \quad \forall i, i'$
  Update $\mathbf{A}$: $\mathbf{A}_{ii'jj'} \leftarrow \mathbf{A}_{ii'jj'} \cdot P\left(u\left(c_{ii'}\right)\right) \cdot P\left(u\left(c_{jj'}\right)\right)$
end while

Chapter 4

Experimental Results

In this chapter we present some experimental results: First, we test the SCMF decomposition scheme proposed in this paper as a multi-assignment graph matching approach. We verify the scheme and demonstrate its advantage, in comparison to the standard Symmetric NMF scheme. We then extensively test our full algorithm, MAGMA, on several synthetic generated scenarios as well as on sequences of real images along with ground truth in increasingly growing difficulty level. We compare the proposed algorithm against a wide variety of state-of-the-art algorithms from both the school of epipolar geometry robust estimation and of graph matching. The robust estimation algorithms tested are: 8-point-RANSAC [22], MLESAC [8], MAPSAC [9], NAPSAC [10], BEEM [12] and BLOGS [13] (the latter two were not tested against the synthetic scenarios, since they require additional appearance parameters related to the SIFT descriptors [1] which were not synthesized); The pairwise GM algorithms tested are: SM [15], GA [14], IPFP [16] and RRWM [17]; The triplets GM algorithms tested are: Zass and Shashua's HyperGraph Matching (HGM) [18], Duchenne's et al. Tensor-based high order Graph Matching (TGM) [20] and Chertok and Keller's Tensor High Order Assignment (THOA) [19].

We have used MLESAC, MAPSAC and NAPSAC implementations from the "structure and motion toolkit" by Philip Torr available online. The BEEM, BLOGS, RRWM and TGM implementations are from the authors' websites. The HGM implementation was given to us by Zass, and RANSAC, SM, GA and THOA were implemented by us. For fair comparison, in the robust estimation algorithms, only the best candidate for each point was considered as input, because those algorithms cannot cope with high outlier rates, and checking more than one candidate for each point will considerably increase the outlier rate.

Both the real and simulated benchmarks allow us to quantitatively evaluate the performances, in terms of accuracy and robustness in various deformations and missing data scenarios, and to fairly compare the algorithms to one another.

4.1 Multi-Assignment Graph Matching Evaluation

We experimentally verified the proposed SCMF multi-assignment scheme by applying it to simulated data, and compared it to the Symmetric NMF scheme. To this end, we conducted the following simulation: We assembled a synthetic but realistic scenario of three planes, each containing 30 uniformly distributed points, captured by two pinhole-model cameras set at different locations, and forming a wide-baseline stereo pair. We then constructed a graph from those image points with additional 9 neighbors per point (total of 10 nearest neighbors per point), and added some random noise and some outliers (points with no correct neighbor). We factorized the resulting graph using each of the methods, and discretized the results to get a hard assignment from each. The process was repeated for 100 times, and the results were saved and compared to the ground-truth. The average ROC (Receiver Operating Characteristic) curves of both methods are displayed in Fig. 4.1. The results are quite convincing in favor of the SCMF scheme for factorization and for the registration of wide-baseline stereo image pairs.

4.2 Matching Synthetic Datasets Using MAGMA

We tested the MAGMA approach presented in Sec. 3.4, by first conducting a set of synthetic experiments. Those experiments were aimed to test different challenging aspects of point matching scenarios. For each trail we constructed a setup of two pinhole-model cameras capturing a scene composed of several groups of 3D points laying in planes at different depthes from the cameras. Each point in the first image was attributed to 10 nearest neighbors in the second image (one of them being the correct match), the points in each image were perturbed by random noise and some outliers (points without any correct neighbor) were induced. At each experiment we held most parameters of the above setup constant while changing one particular parameter at a time: increasing outlier ratio, decreasing the ratio of points having their correct match as the nearest neighbor, increasing noise level, changing the ratio of cameras' focal lengths, in-

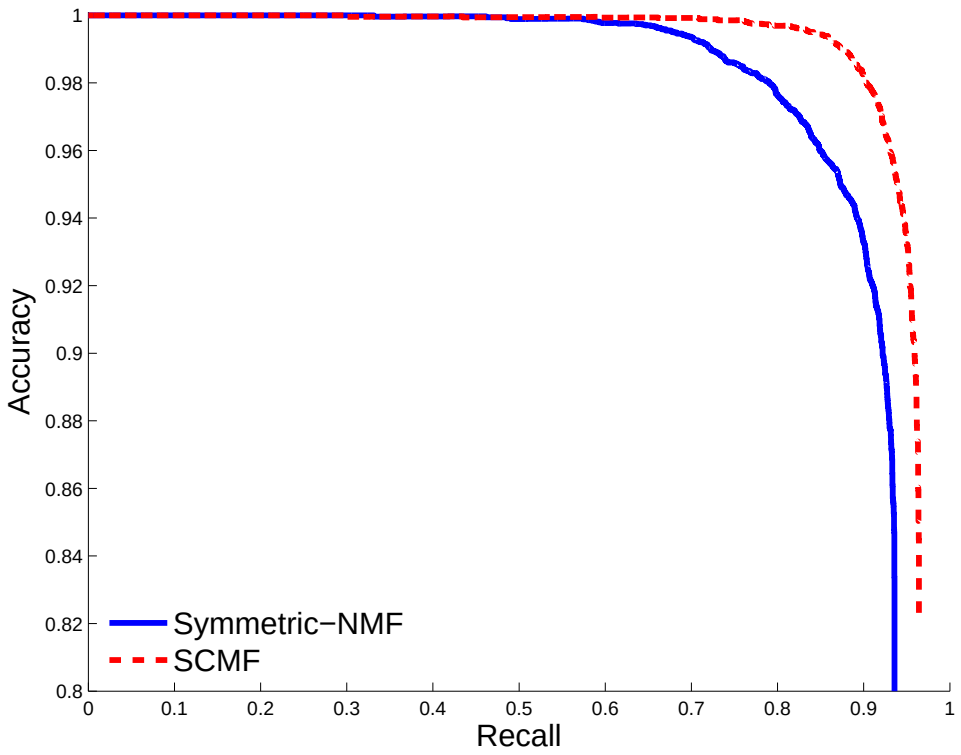


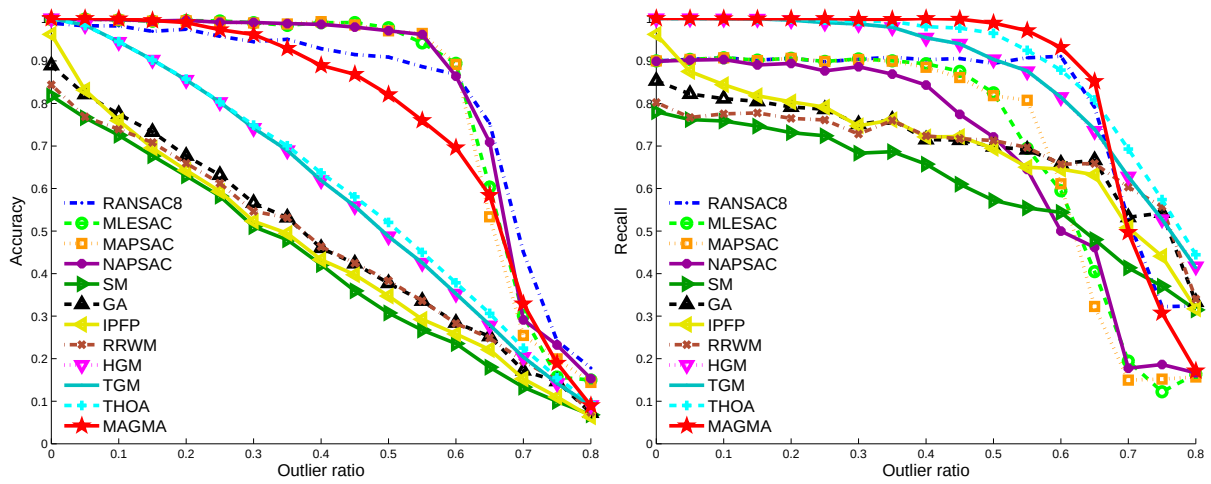
Figure 4.1: ROC curves of the matching simulation results, using the Symmetric-NMF scheme and the SCMF scheme.

creasing the baseline angle and increasing number of planes (or depth variability) in the scene. Other than the tested parameter in each experiment, the rest of the parameters were set to default values according to the following: The synthetic setups were constructed by two cameras with baseline angle of $20°$, both at the same depth from the scene but with different focal-length ($f_2 = 1.1 \cdot f_1$). The scene itself was constructed by total of 120 points in 3 planes, with $1f_1$ depth difference from one another and a maximum distance of $5f_1$ from camera's plane. The default outlier ratio was set to 20%, the ratio of inlier points having their correct match as the nearest neighbor was set to 10% and the noise standard deviation was set to 1 pixel.

The threshold parameter in all robust estimation algorithms (including the MAGMA's RANSAC threshold) was fixed to the same value, $t = 0.01$ (we made sure to compensate for the different pixel normalization in Torr's implementations), the graphs for pairwise GM algorithms were all generated the same way using the same kernel width, $\varepsilon = 25$ (see Eq. 2.5), and the graphs for triplets GM algorithms were all generated the same way using the same kernel width, $\varepsilon = \pi/60$ (see Eq. 2.10), and a full sampling ratio.

We measured the **accuracy** and **recall** percentage of every algorithm in each trail, and re-

peated the process for 20 times, averaging the results for consistency. To fully understand the relative performance of the tested algorithms, both accuracy and recall graphs should be simultaneously considered. The results are presented in Figs. 4.2-4.7.

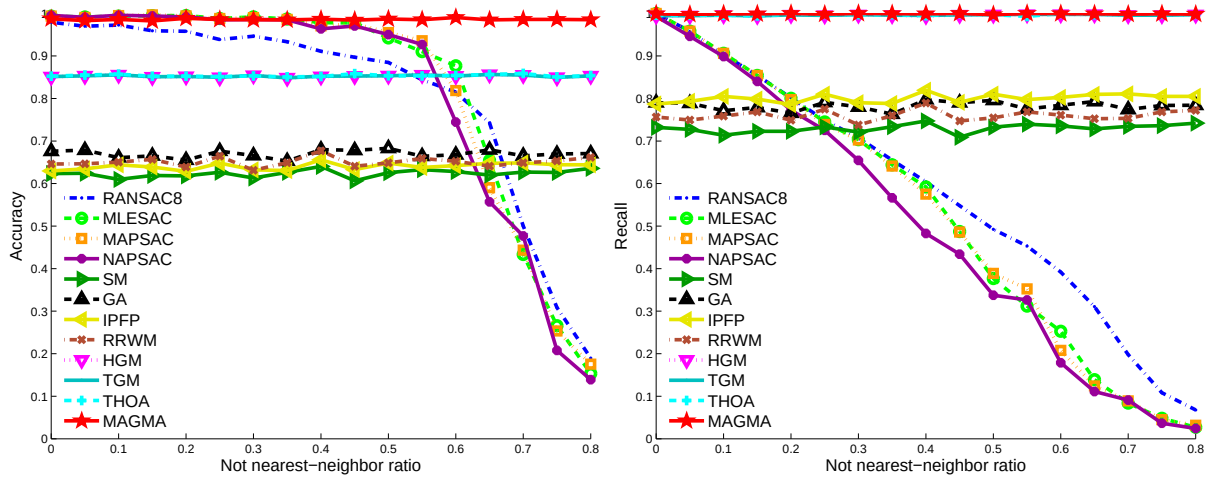


(a) (b)

Figure 4.2: Synthetic test - outlier experiment. (a) Accuracy percentage. (b) Recall percentage.

Figure 4.2 shows the performances of the different algorithms as the outlier percentage is varied from 0% to 80%. The pairwise GM algorithms degrade the fastest as the outlier rate rises, because of its inability to handle scale change between the images, stemming from the focal length change between the two cameras. For up to 25% outliers, MAGMA scores close to perfect accuracy and recall. For more than 25% of outliers, MAGMA's accuracy performance starts degrading while the robust estimation algorithms achieve better accuracy. However, the robust estimation algorithms are significantly inferior with respect to recall performances, in all outlier percentages. The reason for that lies with the fact that not all inliers in the simulation are nearest neighbors, and as stated, one major drawback of robust estimation algorithms is their inability to cope with more than one nearest neighbor.

Figure 4.3 further illustrate this problem. In this simulation all that changes between the trials is the percentage of inliers which are tagged as the nearest neighbors for each point in the source image. All graph based algorithms maintain constant performances in both accuracy and recall percentages, while, as expected the robust estimation algorithms' performances degrade with the increase of the ratio of points not having their correct match as the nearest neighbor. For these algorithms, the recall percentage degrade linearly. Their accuracy degrade as well,



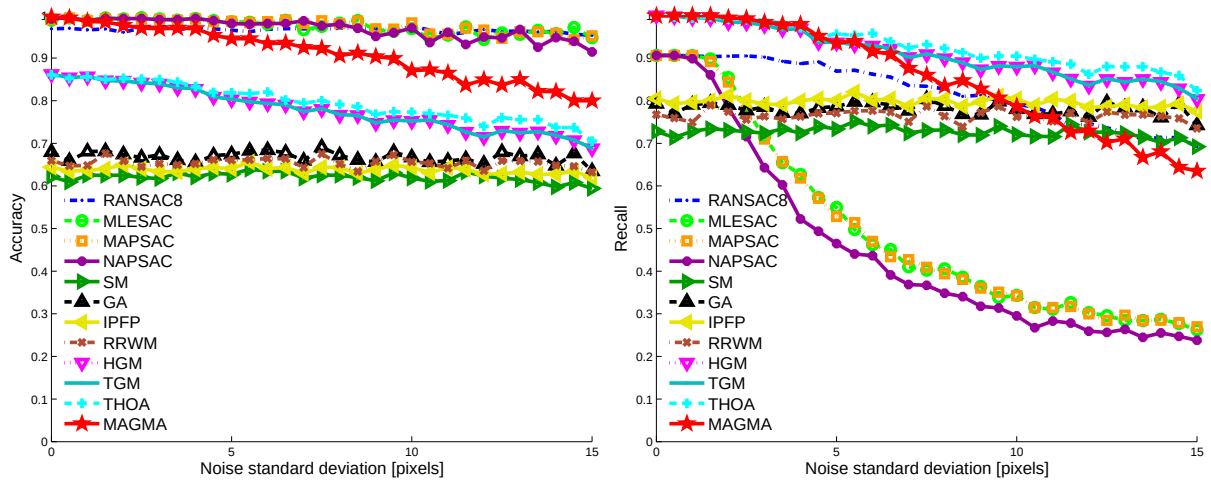
(a) (b)

Figure 4.3: Synthetic test - not-nearest-neighbor experiment. (a) Accuracy percentage. (b) Recall percentage.

as the inlier ratio within the nearest neighbors drops and becomes insufficient for the formation of a strong consensus group. The same is not valid for GM methods (including MAGMA) which are not affected by this parameter at all. In fact, in this experiment MAGMA achieves close-to-perfect scores in all setups.

Figure 4.4 shows what happens as points' localization degrades. In this experiment we perturbed the points' position by Gaussian noise with increasingly growing standard deviation. It is shown again, that though the robust estimation algorithms achieve better accuracy, they are not true competitors to MAGMA when it comes to recall percentage. Also, triplets GM algorithms achieve better recall, while at lower accuracy.

Figure 4.5 illustrate the pairwise GM's weakness of scale change intolerance. In this experiment we changed the ratio of the cameras' focal lengths. This change is equivalent to a scale change between the images, and thus affects the pairwise GM algorithms performances. The pairwise graph is constructed from an isometry affinity measure, so pairwise GM algorithms can only handle isometry transformations (rotations and translations), and are sensitive to scale changes. Hance, those algorithms achieve their optimal performance level when the focal length is the same (no scale change), and their performance drops rapidly with the change in this parameter. Other algorithms remain indifferent to this change. MAGMA performs well as expected, achieving close-to-perfect scores and outperform all other tested methods in all



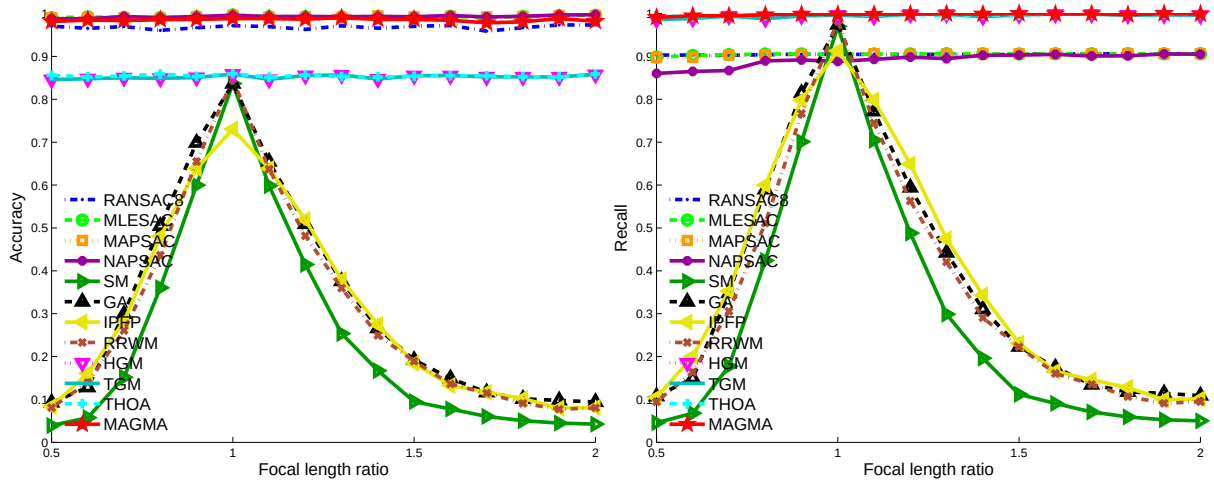
(a) (b)

Figure 4.4: Synthetic test - image noise experiment. (a) Accuracy percentage. (b) Recall percentage.

setups.

Figure 4.6 shows the results of having the baseline of the stereo pair increasing. MAGMA is not affected by this change, as it considers more than one component of the graph, and thus the depth variation and baseline angle do not affect it much. Also not affected are the robust estimation algorithms which are indifferent to such change (as long as the inlier's ratio is the same). Both triplets and pairwise GM degrade much with baseline angle increase due to their one-component restriction.

Finally Fig. 4.7 presents our last synthetic experiment, testing the deviation from the model of MAGMA, by subtracting or adding components to the graph. This effect is achieved by changing the number of planes captured by the cameras from 1 to 5. It can be seen that even though the number of components being calculated by MAGMA in the current setup is constant (based on the SCMF order which was fixed to be $M = 3$), it is still the best algorithm, yielding close to perfect performances for all cases. The robustness of MAGMA to this parameter is due to its probabilistic derivation allowing for soft partitioning of the graph, and the simultaneous discretization of all components, which assures each point correspondence is taken according to the most appropriate assignment group.



(a) (b)

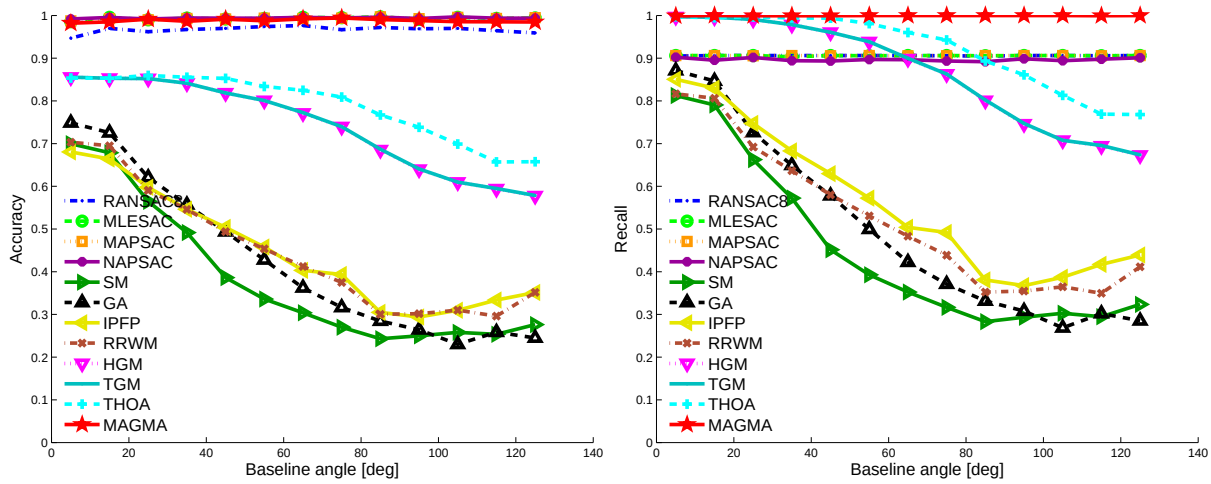
Figure 4.5: Synthetic test - focal length ratio experiment. (a) Accuracy percentage. (b) Recall percentage.

4.3 Matching Real Image Sequences Using MAGMA

We tested the MAGMA approach presented in Sec. 3.4 by applying it to sequences of real images. In these experiments we used two image sequences datasets: fountain-P11 and Herz-Jesu-K7 datasets [37]. Each of the datasets consists, other than the images, of a dense accurate 3D model of the scene, measured by a LIDAR Light Detection and Ranging) sensor, and well calibrated models for the cameras, including poses and directions. Several images from each sequence are presented in Fig. 4.8. The unique setup allows us to extract ground truth information for each point in the scene as to its whereabouts in each image, and thus, to conduct a large-scale quantitative evaluation of all tested approaches.

The experiments exemplify the matching performances in a real-world natural setup. It should be noted that the image pairs are not even remotely compliant with the ideal MAGMA model of several planes or objects of constant depth from the cameras.

We tested each sequence by examining all possible image pairs. For each pair 1500 SIFT points where selected according to their distinctiveness (closest to second closest descriptor distance ratio), along with 2 nearest neighbors per each. The sets of putative points and neighbors were sent to each of the algorithms (except for BLOGS, which calculate its own nearest neighbor differently, and so the nearest neighbors of the same 1500 SIFT points were calculated



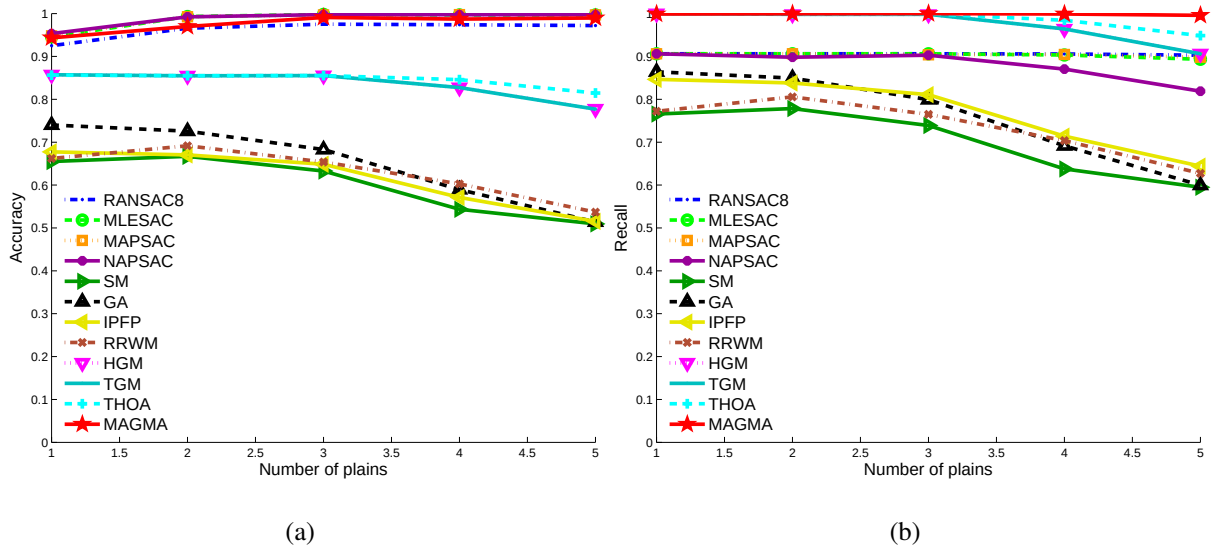
(a) (b)

Figure 4.6: Synthetic test - baseline angle experiment. (a) Accuracy percentage. (b) Recall percentage.

separately by it).

The algorithms' parameters were set as follows: The BEEM and BLOGS parameters were used as given by the authors (which generally produced good results), the threshold parameter in all other robust estimation algorithms was fixed to the same value, $t = 0.01$ (we made sure to compensate for the different pixel normalization in Torr's implementations), the graphs for pairwise GM algorithms were all generated the same way using the same kernel width, $\varepsilon = 5$ (see Eq. 2.5), and the graphs for triplets GM algorithms were all generated the same way using the same kernel width, $\varepsilon = \pi/60$ (see Eq. 2.10), and a full sampling ratio, with the exception of TGM graph, which for computational considerations was generated using a sampling ratio of 0.1.

The results were compared to ground truth. We then measured the **accuracy** and **recall** percentage of every algorithm in each trail. The image pairs are ranging from narrow to wide-baseline, and the results are displayed accordingly in Figs. 4.9 and 4.10. The "Baseline diff" axes is the gap between the images (e.g. $Baseline\ diff = 5$ stands for the average of all pairs 1-6, 6-1, 2-7, 7-2, 3-8, etc.). As the baseline difference grows, some of the parameters examined in the synthetic scenarios in Sec. 4.2 are degraded, affecting matching performance: Along with the increasing of the baseline angle, the motion model becomes more complex, making the assumption of one strong component obeying a single rigid transformation less valid. Also the image appearance varies, reducing features' similarity, and by so doing increasing outlier ratio,



(a) (b)

Figure 4.7: Synthetic test - number of planes experiment. (a) Accuracy percentage. (b) Recall percentage.

decreasing the ratio of points having their correct match as the nearest neighbor and degrading the features' localization (which is equivalent to increasing the noise level).

The results for both datasets validate our approach. Even though the image pairs are not compliant with the MAGMA model, the results still verify it. MAGMA outperforms other GM algorithms, yielding much better accuracy than any of them (including the triplets graph matching schemes). For high values of baseline difference, GM algorithms yield better recall percentage, but their accuracy is so small, indicating that their matching results are meaningless. The inadequate GM performances are due to their one-component assumption, which is a faulty assumption in these cases, and due to the reduction in inlier percentage as the baseline difference grows.

Most robust estimation schemes report accuracies similar to MAGMA for low baseline difference. However, even then, the recall is much smaller in comparison to our approach. The only comparable algorithms for these setups are BEEM and BLOGS. Each of them achieves similar or better accuracies than MAGMA's. In the Herz-Jesu-K7 dataset experiment (Fig. 4.10) MAGMA achieves better recall percentages than both BEEM and BLOGS, and in the fountain-P11 dataset experiment (Fig. 4.9) BLOGS outperform MAGMA for high baseline differences. This can be due to the fact that BLOGS calculate its own initial putative set of correspondences, which could be, by chance better than the initial set used by other algorithms in this

(a)

(b)

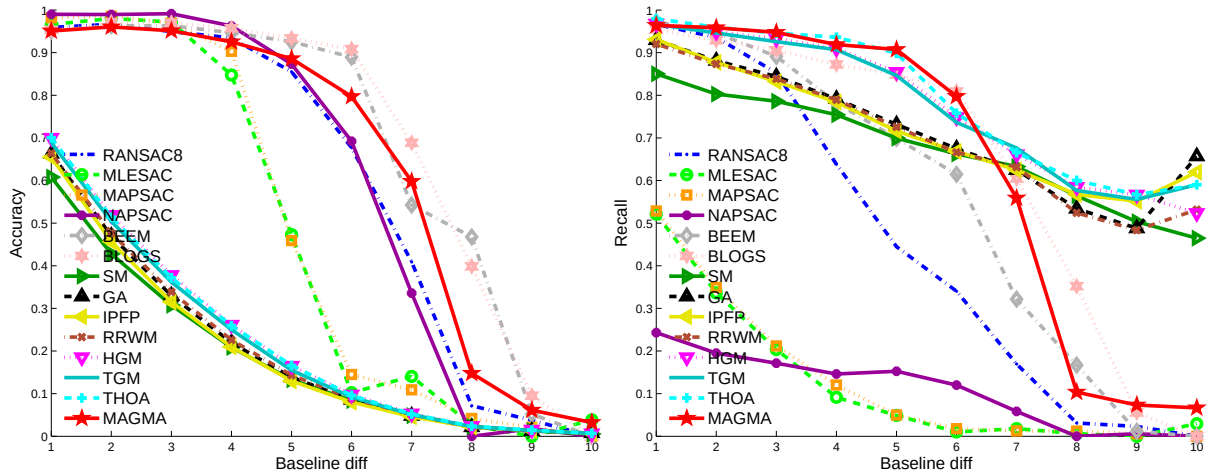
Figure 4.8: Several images from the fountain-P11 (a) and the Herz-Jesu-K7 (b) datasets. The two datasets were used as a benchmark for testing the different algorithms in a sound way.

case. Another reason both advanced robust algorithms perform so well, is that the radiometric acquisition conditions of the captured scenes are quite stable throughout each sequence, making the SIFT descriptors' assignment quality pretty good, so that the correct matches amongst the SIFT neighbors are mostly the nearest ones. This cause neutralizes one of GM's greatest advantages over robust estimation approaches, which is its ability to consider more than one candidate assignment for each point. Still, MAGMA performs well, yielding results comparable to state-of-the-art robust estimation algorithms, and in some cases achieves better results (e.g. in the Herz-Jesu-K7 dataset experiment, Fig. 4.10).

4.4 Conclusion

In this chapter we presented an experimental study of the contributions discussed in this paper. First, it was shown, by conducting a quantitative synthetic simulation, that the proposed SCMF scheme for estimating several disjoint parts in a graph, performs better than the suggested alternative, Symmetric NMF. This result further justifies the contribution of imposing the stochastic constraints in the optimization process.

Next, we comprehensively tested the proposed image matching algorithm, MAGMA, on several simulated setups, each emphasize different challenging aspect of point matching problems. MAGMA achieved outstanding performances in all setups in comparison to other state-

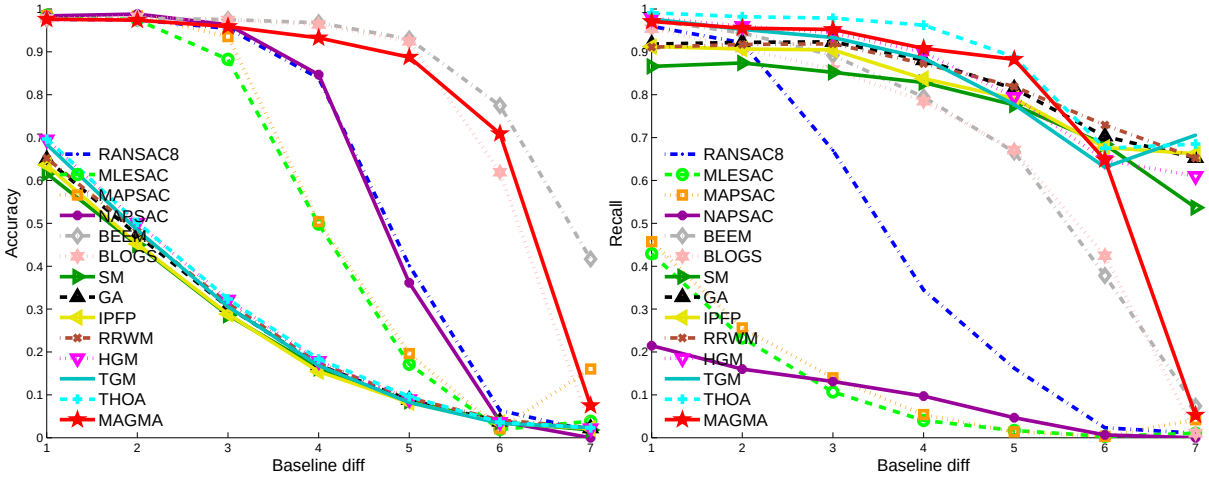


(a) (b)

Figure 4.9: The fountain-P11 sequence experiment. (a) Accuracy percentage. (b) Recall percentage.

of-the-art algorithms.

Finally, we conducted a set of experiments on real image sequences, showing that MAGMA obtains satisfying results, outperforming other GM approaches, and achieves results comparable to the state-of-the-art robust estimation algorithms.



(a) (b)

Figure 4.10: The Herz-Jesu-K7 sequence experiment. (a) Accuracy percentage. (b) Recall percentage.

Chapter 5

Summary and Discussion

5.1 Summary

In this work we addressed the image matching problem, and in particular, the wide-baseline stereo matching problem. State-of-the-art approaches for matching two images were presented, and their drawbacks were discussed. Those include methods for robust estimation of epipolar constraint while pruning outlier matches, and graph based methods for inferring consistent component in an affinity graph or hypergraph, encoding corresponding geometrical relations between small groups of points. All existing algorithms have difficulty treating the wide-baseline setup, due to local and global perspective deformations, affecting both appearance coherence (and thus, reducing inlier rate) and global motion model.

We presented a general probabilistic framework for solving the stereo matching problem. Our Multi-Assignment Graph Matching Algorithm (MAGMA) generalizes the graph matching paradigm, making it more adequate for dealing with wide-baseline scenarios, and more robust to erroneous alignments.

We proposed relaxing the demand for one strongly-connected cluster in the affinity graph, and replacing it with a more appropriate model of several unconnected components. This multi-assignment model was inspired by natural scenes composed of several objects and background, at different depths from the cameras, which are often disjoint by the perspective transformation between the stereo pair. In fact this multi-component situation is the main cause for the failure of ordinary GM methods in matching wide-baseline stereo images. The multi-assignment solution was implemented by an efficient projected Newton optimization scheme, that enforces

stochastic constraints in the process of the solution, and thus, results in a maximum-likelihood estimation of the multiple-assignments structure folded within the graph.

In order to further enhance the robustness of the algorithm to clutter and attenuate false assignments unrelated to the camera's transformation (such as those caused by symmetry or repetitive patterns), we simultaneously assemble the different assignment solutions, and discretize them according to each corresponding point's probability estimate, to form a global set of correspondences. Then, this global assignment is used for robustly estimating the epipolar geometry, binding all the valid assignments. Finally, the unary probabilities inferred by the epipolar geometry are incorporated within the affinity graph, to establish a more accurate probabilistic graph, representing both the heuristic pairwise (or marginalized triplets) probabilities and the epipolar geometry probabilistic prior. The whole process iterates until convergence.

The proposed approach preserves the strengths of both existing paradigms, while somewhat diminishing their weaknesses: The GM strengths, of efficient and robust matching a consistent set of points, is preserved, while being generalized to fit a more generic, multi-component model, adequate to the wide-baseline matching scenario, and the robust epipolar geometry estimation advantage of effectively pruning false correspondences while estimating the fundamental matrix is also utilized, while improving the initial guess of putative assignment by using the output of GM, and inducing additional heuristic constraints restricting the matches to reside in a specific locations along the epipolar lines (which helps filtering out false matches that accidentally agree with the epipolar lines).

An extensive experimental study demonstrated the applicability of the proposed algorithm in robustly matching a large variety of challenging stereo image pairs, including wide-baseline stereo pairs. We have tested our approach in comparison to many state-of-the-art algorithms of both schools of robust estimation and graph matching, via a set of synthetic simulations and large-scale real image tests. Our approach is shown to outperform existing graph matching and epipolar geometry robust estimation state-of-the-art algorithms in many cases, and perform at the top level in all cases. It may also be concluded that although our multi-assignment model presumes a specific number of approximately planner surfaces in the scene, undergoing rigid transformations, the algorithm is not very sensitive to the actual number of such planes, nor to the presence of distinct planes at all, as it still achieves very good performances in real-world setups not complying with the above model.

5.2 Future Research

The work presented in this thesis can be further extended in a number of interesting directions:

1. In our implementation, the number of components in the affinity graph, representing different unconnected parts in the scene, is currently set to a fixed value. Though the experimental results suggest that this constant is best set to a small value such as 3, in complicated scenes where there are many layers of different depths, or alternatively, when the scene is composed of only one or two planar surfaces, it might be better to set it to a different value. A possible generic solution is to adaptively estimate the number of components from the data. There is a vast research literature on the topic of inferring model order from data, from classical Bayesian Information Criterion (BIC) [38], to more modern approaches (e.g. a work by Schmidt et al. [39] proposing to empirically estimate the marginal likelihood of the factorization by Gibbs sampler). Such approaches should be investigated for the purpose of generalizing the suggested solution.

2. One possible way of improving registration results is by using a more generic model for each component's transformation. In our implementation we used the marginalized triplets affinity matrix, which results in a similarity-invariant (or scale-invariant) transformations. However, a 4th order graph could be constructed to facilitate matching affine transformations (or even 5th order graph for homographies - plane-perspective transformations). this model, though much less efficient, might prove worthy in matching some natural setups.

3. In a similar context, the marginalization we applied following the Tensor High Order Assignment approach of Chertock et al. [19], was essential for the rest of our scheme, as we first produced a matrix encoding the high order affinities, and then calculated its RMA using the SCMF scheme. Duchenne et al. [19] suggested using the tensor as a whole, by extending the power iteration used for the calculation of the leading eigenvector in SM [15] to higher dimensions. Similarly, it might be interesting to extend the factorization utilized in our scheme (the SCMF) to higher dimensions, by resorting to schemes similar to the nonnegative tensor factorization [32], but also incorporate stochastic constraints. This way, we may use the high order affinity tensor as a whole, without having

to marginalize it to a matrix form first.

4. Probabilistic priors other than the epipolar prior used in this thesis, may be fused with the affinity matrix (or tensor), as seen in Sec. 3.3, to further improve the overall performance, or to generalize the approach, beyond stereo pair matching, e.g. for matching faces or non-rigid objects. All needs to be done, is to multiply each entry of the affinity matrix by the appropriate unary probabilities, prior to its factorization (or eigendecomposition).

5. The factorization scheme proposed in this thesis uses a least squares formalism, which optimizes the L_2-norm of differences (see Eq. 3.3). Indeed, there are other distance measures which could be optimized for the same purpose, yielding an objective function of a more probabilistic nature, rather than the simple Euclidian one we have used. There is a large number of such distances (or divergences), known to have a probabilistic interpretation. The most famous one is the Kullback-Leibler divergence, which minimization results in an optimal relative-entropy, given that the matrices to be approximated represent normalized probability distributions [30].

6. This work can be generalized and used in other computer vision applications. One obvious such application is symmetry analysis. The SSA algorithm [29] presented in Sec. 2.3.4, is using generalized form of SM for finding self-alignments in an image. As seen in Sec. 3.2, the spectral decomposition framework defined by SSA is not best suited for the job. We argue that due to the probabilistic interpretation and the nonnegativity of the SCMF scheme, it is a much more appropriate instrument for this case. Appearance based priors or reflection/rotation transformation priors can also be applied to the symmetry analysis scheme, similarly to the use of epipolar geometry prior in MAGMA.

7. Due to the framework generality, non-related applications from other scientific fields could also benefit. For example, in computational biology, it is interesting to find relations between semi rigid proteins, whose structure is known through X-ray crystallography. One such example is when two proteins are known to share almost identical composition and\or structure, but have different functionality. In that case we would like to find out what is the difference between the two structures by finding alignments between them. There could be several distinct alignments with different spatial transformation between

them, which we might infer by using SCMF scheme (additional priors can be added, of course, in a style resembling MAGMA).

Other examples from this field include finding the spatial (sometimes subtle) difference between a healthy regular protein and a mutated version of it, and similarly, finding the relation between laboratory crystallized protein's structure and its corresponding structure under different biological conditions (those could be obtained e.g. via a Molecular Dynamics simulation).

8. Just like NMF, the SCMF scheme proposed in this thesis is a general decomposition algorithm which could be utilized for a large variety of applications including clustering, classification and blind source separation tasks. For tasks previously implemented by NMF, whenever a probabilistic set of constraints could be applied, one should consider using SCMF instead of NMF. In our thesis it was demonstrated (Sec. 4.1) that the benefit of enforcing the stochastic constraints in the course of optimization is not negligible. Thus, a myriad of applications could use SCMF optimization scheme to improve their performances. Among such applications are: music analysis and decomposition, Hyperspectral Imaging (HSI), fluorescence information extraction, EEG (electroencephalography) or MEG (magnetoencephalography) analysis, handwriting recognition and many others.

Bibliography

[1] D. G. Lowe, "Object recognition from local scale-invariant features," in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, vol. 2. Los Alamitos, CA, USA: IEEE, August 1999, pp. 1150–1157 vol.2. [Online]. Available: http://dx.doi.org/10.1109/iccv.1999.790410

[2] C. Harris and M. Stephens, *A combined corner and edge detector*. Manchester, UK, 1988, vol. 15, pp. 147–151. [Online]. Available: http://www.cis.rit.edu/~cnspci/references/dip/harris1988.pdf

[3] D. Lowe, "Distinctive image features from scale-invariant keypoints," in *International Journal of Computer Vision*, vol. 20, 2003, pp. 91–110. [Online]. Available: citeseer.ist.psu.edu/lowe04distinctive.html

[4] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," in *Proceedings of the British Machine Vision Conference (BMVC)*, London, GB, 2002, pp. 384–393.

[5] Z. Zhang, "Determining the epipolar geometry and its uncertainty: A review," *Int. J. Comput. Vision*, vol. 27, pp. 161–195, April 1998. [Online]. Available: http://dl.acm.org/citation.cfm?id=290091.290094

[6] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, ISBN: 0521540518, 2004.

[7] O. Chum, J. Matas, and J. Kittler, "Locally Optimized RANSAC," in *DAGM-Symposium*, 2003, pp. 236–243. [Online]. Available: http://www.springerlink.com/content/5xx8b4q181pkey4q

[8] P. H. S. Torr and A. Zisserman, "MLESAC: A New Robust Estimator with Application to Estimating Image Geometry," *Computer Vision and Image Understanding*, vol. 78, no. 1, pp. 138–156, April 2000. [Online]. Available: http://dx.doi.org/10.1006/cviu.1999.0832

[9] P. H. S. Torr, "Bayesian model estimation and selection for epipolar geometry and generic manifold fitting," *Int. J. Comput. Vision*, vol. 50, no. 1, pp. 35–61, October 2002. [Online]. Available: http://dx.doi.org/10.1023/A:1020224303087

[10] D. R. Myatt, P. H. S. Torr, S. J. Nasuto, J. M. Bishop, and R. Craddock, "NAPSAC: high noise, high dimensional robust estimation," in *In BMVC02*, vol. 2, 2002, pp. 458–467. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.62.7082

[11] *Matching with PROSAC - progressive sample consensus*, vol. 1, 2005. [Online]. Available: http://dx.doi.org/10.1109/cvpr.2005.221

[12] L. Goshen and I. Shimshoni, "Balanced exploration and exploitation model search for efficient epipolar geometry estimation," in *Proceedings of the 9th European Conference on Computer Vision - Volume Part II*, ser. ECCV'06. Berlin, Heidelberg: Springer-Verlag, 2006, pp. 151–164. [Online]. Available: http://dx.doi.org/10.1007/11744047_12

[13] A. S. Brahmachari and S. Sarkar, "Hop-diffusion monte carlo for epipolar geometry estimation between very wide-baseline images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 755–762, March 2013. [Online]. Available: http://dx.doi.org/10.1109/TPAMI.2012.227

[14] S. Gold and A. Rangarajan, "A graduated assignment algorithm for graph matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 4, pp. 377–388, 1996.

[15] M. Leordeanu and M. Hebert, "A spectral technique for correspondence problems using pairwise constraints," in *International Conference of Computer Vision (ICCV)*, vol. 2, October 2005, pp. 1482 – 1489.

[16] M. Leordeanu, M. Hebert, and R. Sukthankar, "An Integer Projected Fixed Point Method for Graph Matching and MAP Inference," in *Advances in Neural Information Processing*

Systems 22, Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams, and A. Culotta, Eds., 2009, pp. 1114–1122.

[17] M. Cho, J. Lee, and K. M. Lee, "Reweighted random walks for graph matching," in *Proceedings of the 11th European Conference on Computer Vision: Part V*, ser. ECCV'10. Berlin, Heidelberg: Springer-Verlag, 2010, pp. 492–505. [Online]. Available: http://dl.acm.org/citation.cfm?id=1888150.1888189

[18] R. Zass and A. Shashua, "Probabilistic graph and hypergraph matching," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, June 2008, pp. 1–8.

[19] M. Chertok and Y. Keller, "Efficient high order matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, pp. 2205–2215, December 2010. [Online]. Available: http://dx.doi.org/10.1109/TPAMI.2010.51

[20] O. Duchenne, F. Bach, I.-S. Kweon, and J. Ponce, "A tensor-based algorithm for high-order graph matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2383–2395, December 2011. [Online]. Available: http://dx.doi.org/10.1109/TPAMI.2011.110

[21] R. I. Hartley, "In defense of the eight-point algorithm," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 580–593, 1997.

[22] M. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, June 1981.

[23] B. Triggs, "Joint Feature Distributions for Image Correspondence," in *In Proceedings of the 8th International Conference on Computer Vision*, 2001, pp. 201–208. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.32.4886

[24] O. Faugeras, Z. Zhang, C. Zeller, and G. Csurka, "Characterizing the Uncertainty of the Fundamental Matrix," INRIA, Tech. Rep. RR-2560, June 1995. [Online]. Available: http://hal.inria.fr/inria-00074121/en/

[25] F. Sur, N. Noury, M. odile Berger, and I. N. Grand-est, "Computing the uncertainty of the 8 point algorithm for fundamental matrix estimation," in *In Proc. BMVC*, 2008.

[26] S. S. Brandt, "On the probabilistic epipolar geometry," *Image Vision Comput.*, vol. 26, pp. 405–414, March 2008. [Online]. Available: http://dl.acm.org/citation.cfm?id=1324612.1324634

[27] A. Stojanovic and M. Unger, "Robust detection of point correspondences in stereo images," *Acta Polytechnica*, vol. 47, pp. 23–32, 2007.

[28] A. Egozi, Y. Keller, and H. Guterman, "A probabilistic approach to spectral graph matching," *Submitted*, 2009.

[29] M. Chertok and Y. Keller, "Spectral symmetry analysis," *Accepted for publication. IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009.

[30] D. D. Lee and S. H. Seung, "Algorithms for Non-negative Matrix Factorization," in *NIPS*, 2000, pp. 556–562. [Online]. Available: http://citeseer.ist.psu.edu/lee01algorithms.html

[31] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, October 1999. [Online]. Available: http://www.nature.com/nature/journal/v401/n6755/pdf/401788a0.pdf

[32] A. Cichocki, R. Zdunek, and Amari, "Nonnegative Matrix and Tensor Factorization," *IEEE Signal Processing Magazine*, vol. 25, no. 1, pp. 142–145, 2008. [Online]. Available: http://dx.doi.org/10.1109/msp.2008.4408452

[33] F. Shahnaz, M. W. Berry, Pauca, and R. J. Plemmons, "Document clustering using nonnegative matrix factorization," *Information Processing & Management*, vol. 42, no. 2, pp. 373–386, March 2006. [Online]. Available: http://dx.doi.org/10.1016/j.ipm.2004.11.005

[34] W. Xu, X. Liu, and Y. Gong, "Document clustering based on non-negative matrix factorization," in *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, ser. SIGIR '03. New York, NY, USA: ACM, 2003, pp. 267–273. [Online]. Available: http://dx.doi.org/10.1145/860435.860485

[35] T. Virtanen, "Monaural Sound Source Separation by Nonnegative Matrix Factorization With Temporal Continuity and Sparseness Criteria," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, no. 3, pp. 1066–1074, March 2007. [Online]. Available: http://dx.doi.org/10.1109/tasl.2006.885253

[36] C. Ding, X. He, and H. D. Simon, "On the equivalence of nonnegative matrix factorization and spectral clustering," in *in SIAM International Conference on Data Mining*, 2005.

[37] C. Strecha, W. von Hansen, L. Van Gool, P. Fua, and U. Thoennessen, "On benchmarking camera calibration and multi-view stereo for high resolution imagery," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, vol. 0. Los Alamitos, CA, USA: IEEE, June 2008, pp. 1–8. [Online]. Available: http://dx.doi.org/10.1109/cvpr.2008.4587706

[38] G. Schwarz, "Estimating the Dimension of a Model," *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978. [Online]. Available: http://dx.doi.org/10.2307/2958889

[39] M. N. Schmidt, O. Winther, and L. K. Hansen, "Bayesian non-negative matrix factorization," in *Independent Component Analysis and Signal Separation*. Springer, 2009, pp. 540–547.

הגרף – ממודל המניח מקבץ אחד של נקודות קשירות בגרף המהווה רכיב אחד קשיח יחסית בתמונה העובר טרנספורמציה גלובלית פשוטה – למודל בו מספר רכיבים נפרדים, השומרים, כל אחד בפני עצמו, על עקביות גיאומטרית, כך שכל אחד מאופיין בטרנספורמציה שונה בין התמונות. הבעיה החדשה המתקבלת לאור ההנחה הזו היא בעיה של פירוק גרף למספר רכיבים קשירים שאינם מחוברים ביניהם. למטרה זו מפותחת סכמת אופטימיזציה איטרטיבית המטילה את הפתרון בכל שלב לתת-המרחב המתאים לאילוצים הנובעים מהגדרת בעיית ההתאמה. שנית, אנו מציגים דרך לשילוב המידע ההסתברותי הנובע מהאילוץ האפיפולרי בהליך התאמת הגרף, על מנת לשפר את רובאסטיות ההתאמה. ולבסוף, אנו משלבים את שני הצעדים הללו, לקבלת תהליך איטרטיבי יציב המתכנס במהירות. הגישה ההסתברותית היא כללית, כך שניתן להחליף את המידע ההסתברותי הנובע מהאילוץ האפיפולרי בכל מקור מידע בלתי תלוי אחר, או בשילוב של מדדים הנובעים ממקורות מידע שונים, כגון דמיון במראה הנקודות או טרנספורמציה קשיחה כלשהי, במקרה שזו ידועה.

הגישה המוצעת מצליחה לשמר את יתרונות שתי הגישות הקיימות להתאמת נקודות, תוך שהיא ממתנת את חסרונותיהן: מצד אחד מנוצלת היתירות הקיימת בסצנות טבעיות, בניגוד לגישת שערוך רובאסטי של האילוץ האפיפולרי, ומצד שני, המודל המוצע מתאים יותר לטרנספורמציות המתקבלות בין זוג תמונות סטריאו wide-baseline מאשר המודל הפשטני העומד בבסיסם של אלגוריתמי התאמת גרף. בנוסף, השילוב בין הגישות השונות מחזק את שתיהן – הגרף מתאר טוב יותר את הסתברויות ההתאמה, מאחר והוא מכיל גם אילוץ גלובאלי של גיאומטריה אפיפולרית, ושערוך הגיאומטריה האפיפולרית פשוט יותר, כיוון שהוא מאותחל על ידי אלגוריתם התאמת גרף המתבסס על מספר שכנים קרובים אפשריים לכל נקודה.

אלגוריתם MAGMA נבחן ניסויית מול אלגוריתמי התאמת נקודות עדכניים רבים אחרים משתי האסכולות המתחרות – גישת השערוך הרובאסטי של אילוץ גיאומטרי אפיפולרי, וגישת התאמת הגרף (נבחנו אלגוריתמי התאמת זוגות, כמו גם אלגוריתמי התאמת שלשות). הבחינה נערכה באמצעות ביצוע בדיקות מקיפות הן על נתוני סימולציות במגוון תרחישים מאתגרים והן על נתונים אמיתיים. נמצא כי MAGMA מתחרה היטב באלגוריתמים המהווים את חזית הידע בתחום התאמת התמונה, וכן מעפיל עליהם במקרים רבים.

שימור הזוויות בין כל שלוש נקודות). שתי הגישות עושות לרוב גם שימוש בדמיון בין הנקודות (על סמך ווקטור המאפיינים) על מנת לאתחל את תהליך ההתאמה.

בעוד הגישה הראשונה יעילה כאשר ידוע על קיום אילוץ או מודל תנועה פרמטרי וכאשר האתחול על סמך דמיון הנקודות טוב מספיק, הגישה השנייה פחות רגישה לאתחול, מאחר והיא יכולה להשתמש ביותר מניחוש אחד לכל נקודה, ובנוסף יותר גמישה לעיוותים לא פרמטריים (למשל להתאמות לא קשיחות). גישת התאמת הגרף מניחה הנחה יוריסטית של התמרה אחת גלובלית וקשיחה פחות או יותר בין התמונות, כפי שאכן קורה בסצנות בהן הפרשי העומקים בסצנה אינם משמעותיים ביחס למרחק בין נקודות המבט (narrow-baseline). שימוש בהנחה כזו באלגוריתמי התאמת הגרף מגביל את מיקומי ההתאמות בתמונות, ובכך מפשט את תהליך ההתאמה יחסית לסכמות השערוך הרובאסטי של האילוץ האפיפולרי שאינן מניחות דבר על חלקות או רציפות הסצנה. אולם, ההנחה הזו מהווה גם חולשה משמעותית של אלגוריתמי התאמת הגרף, מכיוון שהעקביות הגיאומטרית בין קבוצות ההתאמות אינה מובטחת, בייחוד בהתאמת צמד סטריאו wide-baseline, ועל כן ביצועי אלגוריתמי ההתאמה מבוססי גרף נפגעים עם עליית המרחק בין המצלמות והעיוות הפרספקטיבי בין התמונות. במקרים קשים אלו ביצועי שני סוגי האלגוריתמים מתדרדרים, ולכן נחשבת התאמת תמונה ב- wide-baseline לבעיה לא פתורה בתחום הראייה הממוחשבת.

בתזה זו אנו מציעים גישה חדשנית לפתרון בעיית התאמת הנקודות בצמד סטריאו wide-baseline (חלק מהשיטות והרעיונות המוצעים יכולים גם לשמש במגוון בעיות התאמה אחרות). הגישה המוצעת מבוססת על האבחנה כי למרות שלא ניתן להניח שהטרנספורמציה בין התמונות היא פשוטה וגלובאלית (כפי שמוצע באלגוריתמי התאמת גרף סטנדרטיים), עדיין לא יהיה נכון לטעון שאין אפשרות להניח לגביה דבר (כמו בגישת שערוך האילוץ האפיפולרי). בסצנות טבעיות ישנה יתירות גבוהה אותה ניתן לנצל לתהליך ההתאמה: לרוב הסצנות המצולמות מורכבות ממספר רבדים (או אובייקטים) סופי, כל אחד בעל עומק אחיד בקירוב יחסית למצלמות. כל רובד בסצנה עובר טרנספורמציה פשוטה יחסית בין התמונות. ולכן, ניתן לטעון כי לרוב התמונות קשורות ביניהן על ידי מספר (מועט יחסית) של טרנספורמציות פשוטות.

האלגוריתם המוצע, **MAGMA** (Multi-Assignment Graph Matching Algorithm), הינו אלגוריתם הסתברותי להתאמת נקודות המנצל יתירות זו לצורך מתן מענה רובאסטי להתאמת הנקודות בין צמד תמונות סטריאו wide-baseline. אלגוריתם זה עושה שימוש בהתאמת גרף בחלקים ובאילוצים הסתברותיים מבוססי גיאומטריה אפיפולרית, ובכך מנצל את היתרונות של אלגוריתמי הגרף ללא וויתור על הגמישות של טרנספורמציות פרספקטיביות מורכבות.

לשם כך, ראשית אנו משפרים את ישימות אלגוריתמי התאמת הגרף ומכלילים אותם למקרים בהם העיוותים הפרספקטיביים משמעותיים, באמצעות הכללת מודל התאמת

תקציר

בהינתן שתי תמונות של אותה הסצנה, בעיית הרגיסטרציה, או "התאמת תמונה", היא הבעיה של מציאת מספר רב של נקודות או מאפיינים תואמים בשתי התמונות, כאשר לכל זוג התאמות מקור משותף בעולם (נקודות קשר או התאמות). בעיה זו הינה בעלת חשיבות רבה בתחומי עיבוד התמונה והראיה הממוחשבת, מאחר והיא מהווה מרכיב חשוב באפליקציות רבות, הכוללות בין היתר זיהוי אובייקטים, שיחזור סצנה תלת-ממדית, שיערוך תנועה, גילוי שינויים מצילומי אוויר, מיזוג מידע ממספר סנסורים ואפליקציות מיפוי שונות.

הפתרון המקובל לבעיית הרגיסטרציה כולל שלושה שלבים: בשלב הראשון מאותרים מספר רב של נקודות עניין בשתי התמונות (נקודות אלו יכולות להיות למשל פינות או כתמים בולטים בתמונה). בשלב השני לכל נקודה ניתן וקטור מאפיינים (**descriptor vector**) אשר מבדיל אותו מנקודות אחרות ואמור לא להשתנות (או להשתנות באופן מינימלי) במהלך התמרות ושינויים שעלולים לנבוע מתנאי הרכשה שונים. בשלב השלישי הנקודות שאותרו בתמונה אחת מותאמות לנקודות שאותרו בתמונה השנייה תוך שימוש בווקטור המאפיינים ובמידע מרחבי.

השלב השלישי, שלב התאמת הנקודות, נחשב כמשימה קשה עקב שינויי מראה בין התמונות ושינויים שונים בתנאי ההרכשה (כמו מיקום המצלמה ושעת הצילום). בתזה זו אנו מתמקדים בשלב התאמת הנקודות, ובמיוחד במקרה בו זוג התמונות מהווה צמד סטראו **wide-baseline**, כלומר שהתמונות צולמו מזוויות שונות ומנקודות שונות במרחב, כך שמרחקן זו מזו גדול יחסית למרחקים בין רבדי עומק שונים בסצנה. מקרה זה קשה במיוחד, שכן העיוותים הפרספקטיביים הנוצרים בו בין התמונות עלולים להיות חמורים. לעיוותים הפרספקטיביים השפעה לוקאלית וגלובלית על ההתאמה – הם משנים את מראה נקודות העניין בתמונה בצורה משמעותית, כך שנפגעת היכולת למצוא נקודות קשר על בסיס מאפייני מראה (דמיון ווקטור המאפיינים), ובנוסף מסבכים את מודל התנועה הגלובלי בין התמונות, כך שלא די בהתמרה פשוטה לתיאור התזוזה – לעיתים קיימים מספר אובייקטים או רכיבים בסצנה אשר במעבר מתמונה לתמונה עוברים טרנספורמציות שונות עקב הפרשי עומקים ביניהם והעיוות הנוצר על ידי הפרספקטיבה.

שתי גישות שונות לביצוע שלב התאמת הנקודות התפתחו במקביל במהלך עשרים השנים האחרונות: האחת משערכת בצורה רובאסטית התאמה פרמטרית או אילוץ פרמטרי כלשהו על ההתאמות (דוגמה נפוצה לאילוץ כזה הינה האילוץ האפיפולרי המתקיים בכל צמד סטריאו של מצלמות חריר – pinhole cameras), ובמהלך פעולת השערוך מסננת נקודות חריגות – כאלה שלא מקיימות את המודל או האילוץ. הגישה השנייה עושה שימוש בתוצאות מתורת הגרפים במטרה למצוא עקביות גיאומטרית בין היחסים המרחביים שחלים בקבוצות קטנות של נקודות (לרוב זוגות או שלשות של נקודות) במעבר מתמונה אחת לשנייה (למשל עבור זוגות – שימור המרחקים בין כל זוג נקודות, ועבור שלשות –

תודות

המחקר נעשה בהנחיית פרופ' ישראל כהן מהפקולטה להנדסת חשמל בטכניון ופרופ' יוסי קלר מבית הספר להנדסה באוניברסיטת בר אילן.

אני מודה לטכניון על התמיכה הכספית הנדיבה בהשתלמותי.

מחקר זה נתמך על-ידי הקרן הלאומית למדע (מענק מס' 1130/11).

ברצוני להביע את תודתי למנחיי, פרופ' יוסי קלר ופרופ' ישראל כהן על הנחייתם, הדרכתם ותמיכתם במהלך כל תקופת המחקר הזה.

אני מבקש גם להודות לדובי וינברגר ולד"ר גילי תלם על ההערות המועילות והדיונים הפוריים.

תודה מיוחדת למשפחתי היקרה – הוריי, יעל ודוד, הורי כלתי, אורלי ויוסי, ואחיי, יונתן ואבנר, על עידודם המתמשך ותמיכתם.

ואחרונה חביבה, ברצוני להביע את אהבתי ותודתי הכנה לאשתי האהובה הילה. תודה על הסבלנות, האהבה האינסופית והתמיכה התמידית. בלעדייך כל זה לא היה מתאפשר.

התאמת תמונות באמצעות התאמת גרף הסתברותית מרובת פתרונות

חיבור על מחקר
לשם מילוי חלקי של הדרישות לקבלת תואר מגיסטר למדעים
בהנדסת חשמל

אורי אוקון

הוגש לסנט הטכניון – מכון טכנולוגי לישראל

סיוון תשע"ד חיפה יוני 2014

התאמת תמונות באמצעות התאמת גרף הסתברותית מרובת פתרונות

אורי אוקון