Doctoral School of Computer Science

Institute of Informatics

University of Szeged

Non-Rigid Registration of Visual Objects

Summary of the Ph.D. Thesis

by

Zsolt Sánta

Supervisor:

Dr. Zoltan Kato

Professor

Szeged

2018

Introduction

Computer vision is a field of analyzing and interpreting visual objects by making use of various computational tools. The fundamental aim is to create *automatized* systems for such tasks, to help or, in some cases, replace human interpretation. Historically, the first applications were working with 2D data. Nowadays, with the development of new 3D acquisition techniques, like Lidar, Microsoft Kinect™ or Intel® RealSense™, processing 3D data becomes more and more important.

An important preparatory step in almost every computer vision process is *image registration*. The main goal of this task is to estimate mappings between different observations of the same scene. The most frequent problem is to match the origins and orientations of the main axes of the underlying coordinate systems by finding a rigid alignment between the input. While in many cases the underlying deformation can be efficiently handled by this simple alignment, there are various applications where we have to deal with larger deformations. The state of the art provides efficient methods to cope with these problems, however, there are lots of open questions as well.

The dissertation addresses the Author's research results in multiple areas of image registration. These results provide solutions for estimating parameters of a wide range of transformations consisting linear, perspective and deformable models.

Deformable Registration of 3D Objects

In this section we propose a general framework to solve the non-linear registration problem between 3D objects. In the following, we will focus on surface alignment, where the effect of the physical deformation is only measurable on the surface of the objects. Since the proposed approach relies on geometrical information only, we are dealing with a true binary registration problem here. The proposed framework is rooted in a successful 2D approach presented in [1], however, the extension to 3D is not straightforward: there are various 3D specific details, which should be handled differently. The proposed general framework is able to work with voxel representation [Sánta and Kato, 2012a][Sánta and Kato, 2012b] and with open and closed triangular surfaces [Sánta and Kato, 2018][Sánta and Kato, 2016a][Sánta and Kato, 2013a]. Possible practical applications also will be investigated, like lung surface and face alignment.

Let us formulate the alignment problem: Given a pair of *template* and *observation* objects denoted by $\mathcal{F}_t \subset \mathbb{R}^3$ and $\mathcal{F}_o \subset \mathbb{R}^3$, respectively, we are looking for the aligning transformation φ such that for all $\mathbf{x} \in \mathcal{F}_t$ there exists a $\mathbf{y} \in \mathcal{F}_o$ satisfying the so-called

identity relation [Sánta and Kato, 2018][Sánta and Kato, 2013a]

$$\varphi(\mathbf{x}) = \mathbf{y} \qquad (1)$$

In classical landmark-based approaches, a large number of corresponding landmarks extracted from $\mathcal{F}_t$ and $\mathcal{F}_o$, giving sufficiently many constraints through Equation (1) to find the parameters of the transformation. What can we do if such point correspondences are not available? Let us integrate out individual point pairs in Equation (1) over the foreground domains of the objects yielding the following equation [Sánta and Kato, 2018][Sánta and Kato, 2013a]:

$$\int_{\mathcal{F}_o} \mathbf{y} \, d\mathbf{y} = \int_{\varphi(\mathcal{F}_t)} \mathbf{z} \, d\mathbf{z}. \qquad (2)$$

While in landmark-based approaches each point correspondence will generate a new equation of the form Equation (1); Equation (2) provides only a very limited number of equations (exactly three for $\mathbb{R}^3$)! In order to generate more equations, observe that Equation (1) (hence Equation (2)) remains valid when a non-linear $\omega : \mathbb{R}^3 \to \mathbb{R}$ function is acting on both sides [1]. Thus adopting a set of independent non-linear functions $\{\omega_i\}_{i=1}^{\ell}$ yields a system of ℓ equations [Sánta and Kato, 2018][Sánta and Kato, 2013a]:

$$\int_{\mathcal{F}_o} \omega_i(\mathbf{y}) \, d\mathbf{y} = \int_{\varphi(\mathcal{F}_t)} \omega_i(\mathbf{z}) \, d\mathbf{z} \qquad i = 1, \ldots, \ell. \qquad (3)$$

The parameters of φ estimated directly as a solution of the system of equations. This implies that we will need ℓ to be at least N, where N is the number of parameters of φ. In practice, however, the input objects usually challenged by various types of noise, therefore an overdetermined system is recommended, thus $\ell \geq N$. This system is solved in the least squares sense with an arbitrary iterative optimizer. For the sake of computational efficiency, we use low-order power functions for $\{\omega_i\}_i^{\ell}$ set:

$$\omega_i(\mathbf{x}) = x_1^{n_i} x_2^{m_i} x_3^{o_i}, \qquad (4)$$

where $\{(n_i, m_i, o_i)\}_{i=1}^{\ell} = \{(a, b, c) \mid a + b + c = O\}$ and $O \in \{0, \ldots, O_{max}\}$.

When we are dealing with objects in the 3D Euclidean space, practically, the possible domains are either surfaces or volumetric objects. The most common surface representation is triangular surface mesh, which can define a surface object (open surfaces) or can be the boundary of a volumetric object (closed surfaces). Assuming that the integrals from Equation (3) are surface integrals over triangular meshes, each of them can be

expressed as sums of integrals over the triangles of each mesh:

$$\sum_{o \in O_\triangle} \int_o \omega_i(\mathbf{y}) \, d\mathbf{y} \approx \sum_{\pi \in \varphi(T_\triangle)} \int_\pi \omega_i(\mathbf{z}) \, d\mathbf{z}, \tag{5}$$

where $T_\triangle \approx \mathcal{F}_t$, $O_\triangle \approx \mathcal{F}_o$ and $\varphi(T_\triangle) \approx \varphi(\mathcal{F}_t)$. The whole formula can be estimated linear time w.r.t. the number of triangles by making use of efficient recursive formulas (an exact and an approximate approach).

A similar formalism can be obtained for volumetric integrals over closed triangular surfaces. In this case, the integrals will be sums of integrals over properly generated tetrahedrons:

$$\sum_{o \in O_\triangle} \operatorname{sgn}(\operatorname{vol}(T_o)) \int_{\mathcal{T}_o} \omega_i(\mathbf{y}) \, d\mathbf{y} \approx \sum_{\pi \in \varphi(T_\triangle)} \operatorname{sgn}(\operatorname{vol}(T_\pi)) \int_{\mathcal{T}_\pi} \omega_i(\mathbf{z}) \, d\mathbf{z}, \tag{6}$$

where T_o is tetrahedron generating using the triangle o and the origin, while $\mathcal{T}_o$ is the continuous enclosed by T_o.

Another common representation for volumetric object is voxel representation. Here we assume that the input is defined as a set of identical volumetric elements (or voxels) with 1 as volume. Let us denote the voxel approximation of the *template* and the *observation* by F_t and F_o, respectively. Since voxels are volumetric objects, integration over these domains will lead to a volumetric integral, which can be estimated as a finite sum over the voxels [Sánta and Kato, 2012a][Sánta and Kato, 2012b]. Therefore, the sides of the equations from Equation (3) will become

$$\sum_{\mathbf{Y} \in F_o} \omega_i(\mathbf{Y}) \approx \int_{\mathcal{F}_o} \omega_i(\mathbf{y}) \, d\mathbf{y}, \tag{7}$$

$$\sum_{\mathbf{Z} \in \varphi(F_t)} \omega_i(\mathbf{Z}) \approx \int_{\varphi(\mathcal{F}_t)} \omega_i(\mathbf{z}) \, d\mathbf{z}, \tag{8}$$

where $i = 1, \ldots, \ell$. However, the generation of $\varphi(F_t)$ could be practically ineffective for high resolution images, hence we will use the integral transformation equations instead. The final form of the system using voxel representation will be

$$\sum_{\mathbf{Y} \in F_o} \omega_i(\mathbf{Y}) = \sum_{\mathbf{X} \in F_t} \omega_i(\varphi(\mathbf{X})) |J_\varphi(\mathbf{X})|, \tag{9}$$

where $i = 1, \ldots, \ell$ and J_φ is the Jacobian of φ. While this system could be used with any parametric transformation model [Sánta and Kato, 2012a][Sánta and Kato, 2012b], an efficient numerical scheme can be derived for models, where the product on the right-hand side is polynomial [Sánta and Kato, 2012a], *e.g..* by making use of polynomial

transformations.

The proposed algorithms were tested on large synthetic datasets. For the triangular surface meshes we used thin plate splines (TPS) transformation, while the voxel framework has been tested with TPS and polynomial transformations. The robustness of the methods have been verified against segmentation errors and model overfitting on a synthetically generated dataset. Moreover, we investigated the influence of TPS control point locations on the results. The method compares favorably to two recent registration approaches CPD [2] and GMMREG [3]. Finally, the algorithm achieved promising results in registering lung CT images, brain surfaces and 3D facial scans, which demonstrates the usefulness of the method in real life applications.

Robust Registration of 2D Images

In this section, we will present our results on registration of 2D images. The work has addressed two different problems related to the algebraic framework introduced in the previous section. In the first problem, the method's tolerance against larger segmentation errors has been investigated [Sánta and Kato, 2014].

In the second problem, we dealt with the ambiguity of the inner parts of shapes, when they are registered using non-rigid transformation models with higher degrees of freedom. This issue has been handled by regularization above, while herein we investigate the possibility of using radiometric information within our algebraic framework [Sánta and Kato, 2016b].

Each proposed method has been experimentally validated on synthetic and real datasets. In addition, we have compared them to multiple state of the art approaches developed in the recent years.

Registration Framework

In the following, we will build our methods upon the 2D registration frameworks introduced in [1, 4]. Let us focus on the pair-wise registration of 2D image regions by estimating the transformation between a *template* and an *observation* patch. Let us assume that the points of these regions are denoted by $\mathcal{F}_t \subset \mathbb{R}^2$ and $\mathcal{F}_o \subset \mathbb{R}^2$, respectively. We are looking for the parameters of a $\varphi : \mathbb{R}^2 \to \mathbb{R}^2$ application specific transformation model which aligns the *template* and the *observation*.

Using these notations, the following relation holds for an arbitrary pair of corresponding points $\mathbf{x} = [x_1, x_2] \in \mathcal{F}_t$ and $\mathbf{y} = [y_1, y_2] \in \mathcal{F}_o$:

$$\varphi(\mathbf{x}) = \mathbf{y} \tag{10}$$

While each of these point correspondences give exactly two constraints on the parameters of φ, reliable landmark extraction could be a challenging task in the presence of non-rigid geometric distortions.

Following [1, 4], instead of extracting individual point-correspondences, let us integrate both sides of Equation (10):

$$\int_{\varphi(\mathcal{F}_t)} \mathbf{z} \, d\mathbf{z} = \int_{\mathcal{F}_o} \mathbf{y} \, d\mathbf{y}. \tag{11}$$

In order to avoid generation of the transformed domain, we can apply the integral transformation to Equation (11), similarly to the 3D framework described in the previous section:

$$\int_{\mathcal{F}_t} \varphi(\mathbf{x}) |J_\varphi(\mathbf{x})| \, d\mathbf{x} = \int_{\mathcal{F}_o} \mathbf{y} \, d\mathbf{y}. \tag{12}$$

Recall that the integral transformation involves the Jacobian determinant of φ on the left-hand side of the equation, which describes the relative change induced by the transformation in the area of an infinitesimal region around each point. $|J_\varphi(\mathbf{x})|$ contains the partial derivatives of the transformation.

Since this equation does not give enough constraints on the parameters, we will apply a set of independent non-linear functions $\{\omega_i \mid \omega_i : \mathbb{R}^2 \to \mathbb{R}\}_{i=1}^{\ell}$ to the coordinates of Equation (11) and Equation (12), respectively, as

$$\int_{\varphi(\mathcal{F}_t)} \omega_i(\mathbf{z}) \, d\mathbf{z} = \int_{\mathcal{F}_o} \omega_i(\mathbf{y}) \, d\mathbf{y}, \tag{13}$$

$$\int_{\mathcal{F}_t} \omega_i(\varphi(\mathbf{x})) |J_\varphi(\mathbf{x})| \, d\mathbf{x} = \int_{\mathcal{F}_o} \omega_i(\mathbf{y}) \, d\mathbf{y}, \tag{14}$$

where $i = 1, \ldots, \ell$. Using polynomial ω functions will eventually lead to a polynomial system of equations and according to [1, 4], the framework is quite robust against several types of geometric noise and segmentation error.

Yet, in previous works, the larger segmentation errors (*i.e.* occlusions or disocclusions) are categorized as a trivial case of failure, similarly to most area based methods. In many applications originated from industrial and medical areas, we only have to deal with the physical deformation due to the well-controlled circumstances of image acquisition, therefore this is not a serious issue in such cases. In a less controlled environment (*e.g.* processing images of surveillance systems), however, we always have to deal with different types of noise and occlusions [Sánta and Kato, 2014].

Affine Alignment of Occluded Shapes

In order to use the proposed framework above, let us assume that pursued transformation is defined by a homogeneous affine matrix $\mathbf{A} \in \mathbb{R}^{3\times3}$:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & 1 \end{pmatrix}. \tag{15}$$

Affine transformations are useful in many practical applications when we are dealing with linear or perspective motions. In the latter case, the affine model is used as a first-order approximation of the deformation. This is a convenient way to model the transformation when we work on images taken with classical pin-hole cameras, but the distance of the camera from the objects is large compared to the size of the object [5].

Using such notations, the identity relation from Equation (10) will become [Sánta and Kato, 2014]

$$\mathbf{A}\bar{\mathbf{x}} = \bar{\mathbf{y}} \quad \Leftrightarrow \quad \bar{\mathbf{x}} = \mathbf{A}^{-1}\bar{\mathbf{y}}. \tag{16}$$
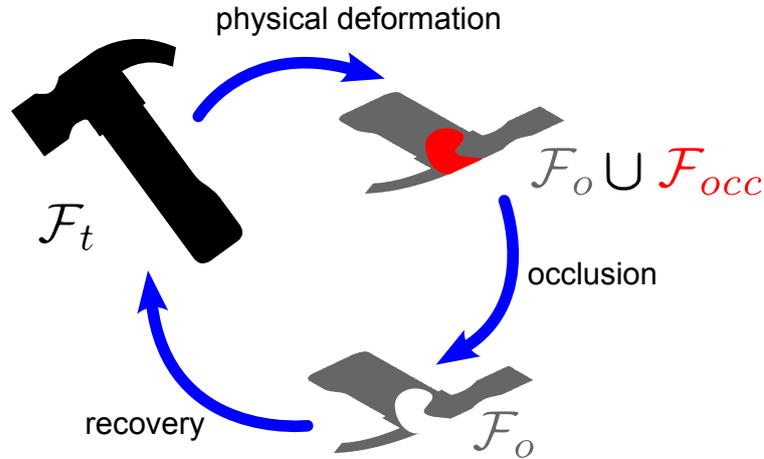

Figure 1: Relation between the *template* and the occluded *observation*.

While this relation is true when the *observation* is a perfect image of the *template*, in the presence of occlusions this will not be valid anymore. Let us denote the *occluded* regions of the *observation* by $\mathcal{F}_{occ}$ (see Figure 1 for an example) [Sánta and Kato, 2014]. Then, we can define an object level identity relation as:

$$\mathbf{A}(\mathcal{F}_t) = \mathcal{F}_o \cup^* \mathcal{F}_{occ} \iff \mathcal{F}_t = \mathbf{A}^{-1}(\mathcal{F}_t \cup^* \mathcal{F}_{occ}), \tag{17}$$

where $\cup^*$ denotes the disjoint union of two sets (*i.e.* there are no overlaps between the

sets). Using these sets as domains in Equation (14) leads to the following system:

$$\int_{\mathbf{A}(\mathcal{F}_t)} \omega_i(\bar{\mathbf{z}}) \, d\bar{\mathbf{z}} = \int_{\mathcal{F}_o \cup^* \mathcal{F}_{occ}} \omega_i(\bar{\mathbf{y}}) \, d\bar{\mathbf{y}} \tag{18}$$

$$\int_{\mathcal{F}_t} \omega_i(\bar{\mathbf{x}}) \, d\bar{\mathbf{x}} = \int_{\mathbf{A}^{-1}(\mathcal{F}_o \cup^* \mathcal{F}_{occ})} \omega_i(\bar{\mathbf{p}}) \, d\bar{\mathbf{p}}, \tag{19}$$

where $i = 1, \ldots, \ell$. For the sake of simplicity, we will assume that the ω_i functions are acting on the Cartesian coordinates of the points [Sánta and Kato, 2014]. Observe that for all non-zero $\mathbf{A}$ affine transformations, the following equation is valid:

$$\mathbf{A}^{-1}(\mathcal{F}_o \cup^* \mathcal{F}_{occ}) = \mathbf{A}^{-1}(\mathcal{F}_o) \cup^* \mathbf{A}^{-1}(\mathcal{F}_{occ}) \tag{20}$$

Substituting this observation into the integration domain of the systems from Equation (18) and Equation (19) and applying the basic properties of Lebesgue-integrals, we get:

$$\int_{\mathbf{A}(\mathcal{F}_t)} \omega_i(\bar{\mathbf{z}}) \, d\bar{\mathbf{z}} = \int_{\mathcal{F}_o} \omega_i(\bar{\mathbf{y}}) \, d\bar{\mathbf{y}} + \int_{\mathcal{F}_{occ}} \omega_i(\bar{\mathbf{y}}') \, d\bar{\mathbf{y}}' \tag{21}$$

$$\int_{\mathcal{F}_t} \omega_i(\bar{\mathbf{x}}) \, d\bar{\mathbf{x}} = \int_{\mathbf{A}^{-1}(\mathcal{F}_o)} \omega_i(\bar{\mathbf{q}}) \, d\bar{\mathbf{q}} + \int_{\mathbf{A}^{-1}(\mathcal{F}_{occ})} \omega_i(\bar{\mathbf{q}}') \, d\bar{\mathbf{q}}' \tag{22}$$

where $i = 1, \ldots, \ell$. Similarly to the affine framework proposed in [4], we use simple power functions for the $\{\omega_i\}_{i=1}^{\ell}$ set and choosing $\ell \geq 6$ yields an overdetermined system, which is then solved in the least-squares sense [Sánta and Kato, 2014].

Within the proposed framework, several algebraically equivalent solutions are possible. For example, a trivial solution is when all parameters of the transformation become zero and the target shape will be the occlusion. In order to overcome this issue, the problem has to be regularized [Sánta and Kato, 2014]. Herein, we use a simple regularization by restricting the size of the occluded areas to be minimal and using Equation (22) yields the following problem:

$$\int_{\mathcal{F}_t} \omega_i(\bar{\mathbf{x}}) \, d\bar{\mathbf{x}} = \int_{\mathbf{A}^{-1}(\mathcal{F}_o)} \omega_i(\bar{\mathbf{q}}) \, d\bar{\mathbf{q}} + \int_{\mathbf{A}^{-1}(\mathcal{F}_{occ})} \omega_i(\bar{\mathbf{q}}') \, d\bar{\mathbf{q}}'$$
$$\text{such that } |\mathbf{A}^{-1}(\mathcal{F}_{occ})| \to \min \qquad i = 1, \ldots, \ell, \tag{23}$$

where $|\mathbf{A}^{-1}(\mathcal{F}_{occ})|$ denotes the area of $\mathbf{A}^{-1}(\mathcal{F}_{occ})$. Note that, for Equation (22) a similar problem can be defined, which will be equivalent to Equation (23).

In order to reduce the computational complexity, we represent the shapes as interior domains of polygons extracted from the their contours [Sánta and Kato, 2014]. Let us denote the polygonal approximations of the borders of $\mathcal{F}_t$, $\mathcal{F}_o$ and $\mathcal{F}_{occ}$ by $T_\Diamond$, $O_\Diamond$ and $M_\Diamond$, respectively. Using these approximations, the integration domain for the occluded
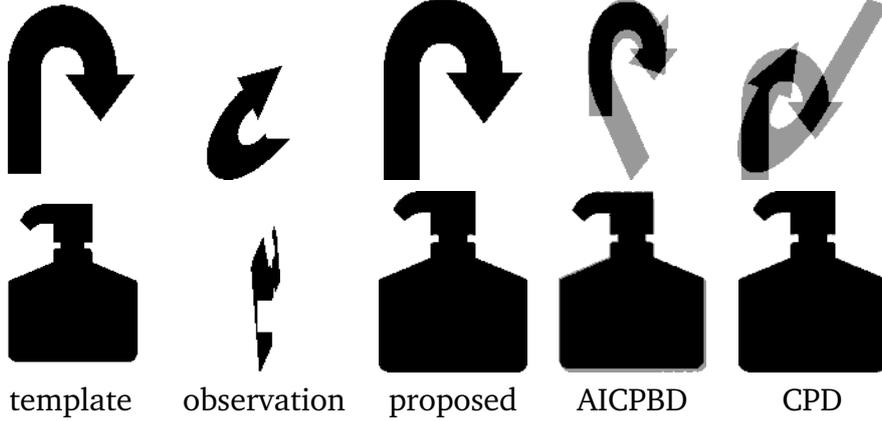
Figure 2: Examples from the synthetic dataset. In the last three columns, we denote the template shape with black and the alignment errors with gray colors.

parts can be determined by making use of simple Boolean operations:

$$\mathbf{A}^{-1}(M_\diamond) = T_\diamond \setminus \mathbf{A}^{-1}(O_\diamond), \tag{24}$$

$$M_\diamond = \mathbf{A}^{-1}(O_\diamond) \setminus \mathbf{A}(T_\diamond), \tag{25}$$

where $\setminus$ denotes the relative complement operator between two polygons. Integrating power functions over polygons in 2D is similarly convenient as integrating over triangular surfaces in 3D. Moreover, the computational complexity is also minimal due to recursive formulas. Therefore the $\{\omega_i\}$ can be defined as follows:

$$\omega_i(\bar{\mathbf{x}}) = x_1^{n_i} x_2^{m_i}, \tag{26}$$

where $\{(n_i, m_i)\}_{i=1}^{l} = \{(a, b) \mid a + b = O\}$ and $O \in \{0, \ldots, O_{max}\}$.

In our experiments, a system of 9 equations has been generated using the above formula, hence $O_{max} = 3$. The algorithm has been implemented in MATLAB (with MEX routines for determining the occluded areas). In order to quantitatively evaluate the performance of the proposed method, we have tested the algorithm on 3000 synthetically generated shapes obtained from the homepage of [6] (`http://www.inf.u-szeged.hu/~kato/software/affbinregdemo.html`). We applied randomly generated occlusions to each test image and obtained four different dataset with the following amounts of occlusion: $0\%, 10\%, 20\%$ and 30%. We have also compared our results to two recent registration approaches the AICPBD [7] and the CPD [2] methods.

Summarizing the experimental results, we can say that the algorithm works quite efficiently when the underlying deformation is affine or can be approximated by an affine transformation well, as long as the size of occlusion is under 20%.

8

Non-rigid Registration of Covariant Functions

In the second topic of this section, we propose an algorithm for non-rigid registration of image patches using a similar algebraic framework [Sánta and Kato, 2016b]. So far, we were focusing on using only geometric information to solve the registration problem. While in many cases this is sufficient, there are problems where the geometric information is not meaningful enough to give constraints on the solution.

A convenient way to deal with such difficulties is to introduce more prior information to the problem. When we are dealing with 2D images, the simplest idea is to add intensity information, if it is available.
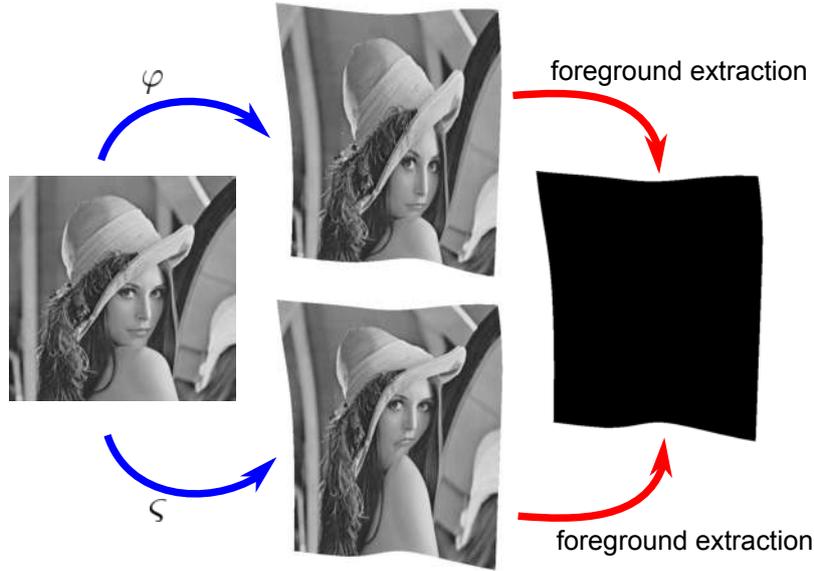

Figure 3: The ambiguity of shape-based registration. In this example, two different deformable transformations applied to the same template give very different observations, but they have the same foreground region after segmentation.

Let us assume that we have two intensity functions for the *template* and the *observation* denoted by $T : \mathcal{F}_t \subset \mathbb{R}^2 \to \mathbb{R}$ and $O : \mathcal{F}_o \subset \mathbb{R}^2 \to \mathbb{R}$, respectively. Assuming that the identity relation from Equation (10) holds and these functions are *covariant* with respect to the φ transformation, we get the following [Sánta and Kato, 2016b]

$$T \circ \varphi = O \tag{27}$$

$$T(\mathbf{x}) = O(\varphi(\mathbf{x})) = O(\mathbf{y}). \tag{28}$$

In the current approach, our aim is to combine the robustness of the geometric approach with the regularizing effect of the intensity functions [Sánta and Kato, 2016b]. This is obtained by multiplying Equation (10) and Equation (28):

$$\int_{\mathcal{F}_t} \varphi(\mathbf{x}) \, T(\mathbf{x}) \, |J_\varphi(\mathbf{x})| \, d\mathbf{x} = \int_{\mathcal{F}_o} \mathbf{y} \, O(\mathbf{y}) \, d\mathbf{y}. \tag{29}$$

Then, similarly to Equation (14), we will apply the $\{\omega_i\}$ set to the image coordinates

$$\int_{\mathcal{F}_t} \omega_i(\varphi(\mathbf{x})) \, T(\mathbf{x}) \, |J_\varphi(\mathbf{x})| \, d\mathbf{x} = \int_{\mathcal{F}_o} \omega_i(\mathbf{y}) \, O(\mathbf{y}) \, d\mathbf{y}, \tag{30}$$

where $i = 1, \ldots, \ell$. This coupled system of equations is then solved in the least-squares sense, providing the parameters of the pursued transformation [Sánta and Kato, 2016b].

Experimental results show the capabilities of the proposed approach both on synthetic and real images. For the former experiments, we generated synthetic datasets for affine and TPS models, by applying randomly generated transformations to a set of *template* images.

The proposed approach has been able to cope with the ambiguity problem [Sánta and Kato, 2016b]. Similarly to the 3D case, we investigated the effect of the control point locations on the outcome for the TPS model. Our experiments have shown that the runtime of the algorithm can be reduced by solving the problem with a coarse-to-fine hierarchical strategy [5]. In our implementation, we use Gaussian image pyramids computed by the algorithm described in [8]. The robustness against intensity noise has been verified as well. The method compares favorably to two recent registration approaches DROP [9] and SPECDEM [10]. We showed the capabilities of the method on two real datasets, too. In the first dataset, we used photos taken from T-shirts containing textured patterns. For the second dataset, we used a public paper-bend set from [11].

Ad-hoc Mobile Camera Network Calibration

In this section, we propose a camera network pose estimation algorithm, which was inspired by the recent works on low-rank rectification. The aim of the proposed method is to localize the network in a 3D scene with respect to a 3D structure [Sánta and Kato, 2013b]. In other words, the origin of the estimated network will be attached to a well-defined 3D location. Our initial assumption excludes the availability of a calibration pattern, thus we also have to reconstruct such structure using the camera images.

The proposed approach has three main steps:

1. Estimation of the relative pose of the cameras within the network w.r.t. an arbitrary *main camera*. For this task, we used the *Normalized 8-point Algorithm* to determine essential matrices and relative poses between each pair of cameras.

2. Computation of the absolute pose between the *main camera* and a 3D structure. This is achieved by making use of the Transform Invariant Low-rank Textures (TILT) [12] approach. This method estimates a planar homography which is able to transform 2D patches to have minimal row or column ranks.

3. Estimation of the relative scale factors of the previous two coordinate systems yielding a fully calibrated network w.r.t. the 3D scene. For this task, we developed a simple one-parameter registration approach, which is built upon the relative poses (from Point 1) and the TILT homography (from Point 2).

The calibration algorithm is designed in a distributed way: each mobile is processing its own image and sends only the minimum amount of data towards other mobiles.

Although, for many steps of the algorithm it is necessarily to pick a *main camera*, the choice of this particular camera is arbitrary and has no major influence on the quality of the final estimates. Robustness and stability of the algorithm have been tested on a large synthetic dataset as well as on real images taken in urban environments.
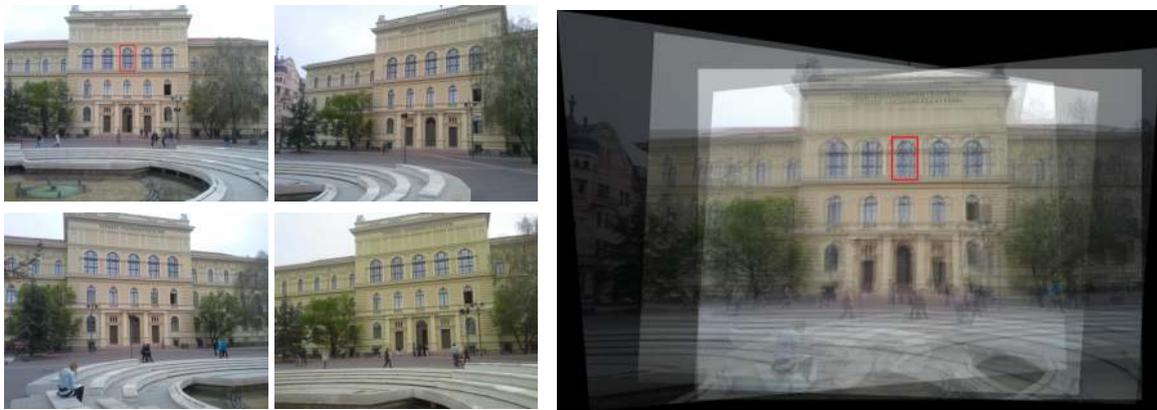
Figure 4: Results on real images. The last image shows the overlaid back projected camera images using the estimated camera matrices.

Summary of the Author's Contributions

In the following, I summarized my results into two main thesis groups. In the first one, I present my findings on registering 3D objects, while in the second one my results on 2D shape registration are shown. In Table 1, the connections between the thesis points and the corresponding publications are displayed.

I.) **Registration of 3D Objects**

Inspired by the already published 2D registration framework [1], the method can be extended for registering 3D objects. The basic idea is to set up a system of non-linear equations whose solution directly provides the parameters of the aligning transformation modeled by a parametric transformation model.

(a) When considering general 3D surfaces as input objects, the basic integrals of the equations will be surface integrals. I derived two recursive numerical

schemes (an exact and an approximate) to efficiently estimate surface integrals over triangular surface meshes. I tested the proposed framework on a large synthetic dataset using thin plate spline (TPS) transformation model. I empirically verified the robustness of the method against segmentation errors and compared the results to two recent registration framework [2, 3]. Finally, I showed practical application of the method for aligning 3D facial scans.

(b) Registering 3D volumetric objects can be realized by making use of volumetric integrals as well. For this purpose, I investigated two object representations: a voxel based and a closed triangular surface based approach. For voxel representation, following the theoretical result from [1], I derived an efficient numerical scheme for estimating the integrals in the case of polynomial transformations. For triangular surface mesh representation, following the general surface approach, I gave an efficient numerical scheme based on tetrahedrons in the case of TPS transformations. I demonstrated the efficiency of the methods on a large synthetic dataset. I made experiments to verify the robustness of the methods against segmentation errors and model overfitting. I showed practical applications of the method for registering lung CT scans and brain surfaces.

II.) **Registration of 2D Shapes**

This thesis group summarize my results on registering binary and grayscale images. In the first topic, I dealt with the alignment of occluded binary shapes, which is a common problem for registering images taken in less controlled environments. In the second topic, I dealt with the ambiguity of the inner parts of shapes, when they are registered using non-rigid transformation models with higher degrees of freedom. Finally, I developed a method for calibrating ad-hoc camera networks.

(a) The affine registration methods for binary shapes published in [4, 6] can be adapted to handle occluded shapes by appropriately choosing the integration domains. In the proposed approach, I represented the shapes as polygons and determined iteratively the occluded areas. Then, using these areas as integration domains, I estimated the best affine transformation between the shapes. I demonstrated the efficiency of the method on a large synthetic dataset. I compared the results to two recent registration methods [2, 7]. I showed practical applicability of the method on images taken in urban environments containing static and dynamic occlusions.

(b) The general framework from [1] can be further regularized by making use of grayscale images. For this purpose, I developed a formalism, where the geometric and intensity information are used in a coupled system of equa-

tions. I verified empirically the efficiency of the method on a large synthetic dataset. I showed experimentally the robustness of the method against additive zero-mean Gaussian noise. I compared the results to two recent registration method [9, 10]. Finally, I showed practical applicability of the framework on real images and on a public dataset.

(c) Recent results on Transform Invariant Low-rank Textures (TILT) [12] can be used for calibrating camera networks. I developed an algorithm which is able to determine the absolute pose of a camera network by making use of planar homographies extracted from TILT features. I verified the efficiency and stability of the proposed method on a large synthetic dataset and real images taken with mobile cameras.

	I		II		
	a	b	a	b	c
[Sánta and Kato, 2012a]		●			
[Sánta and Kato, 2012b]		●			
[Sánta and Kato, 2013a]		●			
[Sánta and Kato, 2016a]	●				
[Sánta and Kato, 2018]	●	●			
[Sánta and Kato, 2014]			●		
[Sánta and Kato, 2016b]				●	
[Sánta and Kato, 2013b]					●

Table 1: The connection between the thesis points and publications.

Publications

Refereed Articles

[Sánta and Kato, 2018]   Z. Sánta and Z. Kato, "Elastic Alignment of Triangular Surface Meshes", *International Journal on Computer Vision*, 2018, Accepted.

Refereed Conference Papers

[Sánta and Kato, 2016a]   Z. Sánta and Z. Kato, "3D Face Alignment without Correspondences", in *Proceedings of the ECCV Workshop on 3D Face Alignment in the Wild*, Amsterdam, Netherlands, Oct. 2016, pp. 521–535. DOI: 10.1007/978-3-319-48881-3_36.

[Sánta and Kato, 2016b]   Z. Sánta and Z. Kato, "An Algebraic Framework for Deformable Image Registration", in *Proceedings of International Conference on Pattern Recognition*, Cancun, Mexico, Dec. 2016, pp. 3792–3797. DOI: 10.1109/ICPR.2016.7900225.

[Tanács *et al.*, 2015]   A. Tanács, A. Majdik, L. Hajder, J. Molnár, Z. Sánta, and Z. Kato, "Collaborative Mobile 3D Reconstruction of Urban Scenes", in *Proceedings of the ACCV Workshop on Intelligent Mobile and Egocentric Vision*, C. V. Jawahar and S. Shan, Eds., ser. Lecture Notes in Computer Science, vol. 9010, Singapore: Springer International Publishing, 2015, pp. 486–501, ISBN: 978-3-319-16633-9. DOI: 10.1007/978-3-319-16634-6_36.

[Sánta and Kato, 2014]   Z. Sánta and Z. Kato, "Affine Alignment of Occluded Shapes", in *Proceedings of International Conference on Pattern Recognition*, IAPR, Stockholm, Sweden: IEEE, Aug. 2014, pp. 2155–2160. DOI: 10.1109/ICPR.2014.375.

[Sánta and Kato, 2013a]   Z. Sánta and Z. Kato, "Correspondence-Less Non-Rigid Registration of Triangular Surface Meshes", in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Portland, Oregon: IEEE, Jun. 2013, pp. 2275–2282. DOI: 10.1109/CVPR.2013.295.

[Sánta and Kato, 2013b]   Z. Sánta and Z. Kato, "Pose Estimation of Ad-hoc Mobile Camera Networks", in *Proceedings of International Conference on Digital Image Computing: Techniques and Applications*, Hobart, Tasmania, Australia: IEEE, Nov. 2013, pp. 88–95. DOI: 10.1109/DICTA.2013.6691514.

[Sánta and Kato, 2012a]   Z. Sánta and Z. Kato, "A Unifying Framework for Non-linear Registration of 3D Objects", in *Proceedings of IEEE International Conference on Cognitive Infocommunications*, Kosice, Slovakia: IEEE, Dec. 2012, pp. 547–552. DOI: 10.1109/CogInfoCom.2012.6422041.

[Sánta and Kato, 2012b]   Z. Sánta and Z. Kato, "Elastic Registration of 3D Deformable Objects", in *Proceedings of International Conference on Digital Image Computing: Techniques and Applications*, Fremantle, Western Australia: IEEE, Dec. 2012. DOI: 10.1109/DICTA.2012.6411674.

References

[1]   C. Domokos, J. Nemeth, and Z. Kato, "Nonlinear Shape Registration without Correspondences", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 5, pp. 943–958, 2012. DOI: 10.1109/TPAMI.2011.200.

[2]   A. Myronenko and X. Song, "Point Set Registration: Coherent Point Drift", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 12, pp. 2262–2275, 2010, ISSN: 0162-8828. DOI: 10.1109/TPAMI.2010.46.

[3]   B. Jian and B. Vemuri, "Robust Point Set Registration Using Gaussian Mixture Models", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1633–1645, Aug. 2011, ISSN: 0162-8828. DOI: 10.1109/TPAMI.2010.223.

[4]   A. Tanács, J. Lindblad, N. Sladoje, and Z. Kato, "Estimation of linear deformations of 2D and 3D fuzzy objects", *Pattern Recognition*, vol. 48, no. 4, pp. 1391–1403, 2015, ISSN: 0031-3203. DOI: 10.1016/j.patcog.2014.10.006.

[5] B. Zitová and J. Flusser, "Image registration methods: A survey", *Image and Vision Computing*, vol. 21, no. 11, pp. 977–1000, 2003, ISSN: 0262-8856. DOI: 10.1016/S0262-8856(03)00137-9.

[6] C. Domokos and Z. Kato, "Parametric Estimation of Affine Deformations of Planar Shapes", *Pattern Recognition*, vol. 43, no. 3, pp. 569–578, 2010. DOI: 10.1016/j.patcog.2009.08.013.

[7] J. Zhu *et al.*, "Robust affine iterative closest point algorithm with bidirectional distance", *Computer Vision, IET*, vol. 6, no. 3, pp. 252–261, 2012, ISSN: 1751-9632. DOI: 10.1049/iet-cvi.2011.0178.

[8] P. Burt and E. Adelson, "The Laplacian Pyramid as a Compact Image Code", *IEEE Transactions on Communications*, vol. 31, no. 4, pp. 532–540, Apr. 1983, ISSN: 0090-6778. DOI: 10.1109/TCOM.1983.1095851.

[9] B. Glocker, A. Sotiras, N. Komodakis, and N. Paragios, "Deformable Medical Image Registration: Setting the State of the Art with Discrete Methods", *Annual Review of Biomedical Engineering*, vol. 13, no. 1, pp. 219–244, 2011. DOI: 10.1146/annurev-bioeng-071910-124649.

[10] H. Lombaert *et al.*, "Spectral Log-Demons: Diffeomorphic Image Registration with Very Large Deformations", *International Journal on Computer Vision*, vol. 107, no. 3, pp. 254–271, 2014. DOI: 10.1007/s11263-013-0681-5.

[11] M. Salzmann, R. Hartley, and P. Fua, "Convex Optimization for Deformable Surface 3-D Tracking", in *Proceedings of International Conference on Computer Vision*, Oct. 2007, pp. 1–8. DOI: 10.1109/ICCV.2007.4409031.

[12] Z. Zhang, A. Ganesh, X. Liang, and Y. Ma, "TILT: Transform Invariant Low-Rank Textures", *International Journal on Computer Vision*, vol. 99, no. 1, pp. 1–24, 2012, ISSN: 0920-5691. DOI: 10.1007/s11263-012-0515-x.