

Dynamic Feature Ordering for Efficient Registration

Tat-Jen Cham

James M. Rehg

Cambridge Research Laboratory
Compaq Computer Corporation
One Kendall Square, Ste 721
Cambridge, MA 02139
{tjc,rehg}@crl.dec.com

Abstract

Existing sequential feature-based registration algorithms involving search typically either select features randomly (eg. the RANSAC[8] approach) or assume a predefined, intuitive ordering for the features (eg. based on size or resolution). This paper presents a formal framework for computing an ordering for features which maximizes search efficiency. Features are ranked according to matching ambiguity measure, and an algorithm is proposed which couples the feature selection with the parameter estimation, resulting in a dynamic feature ordering. The analysis is extended to template features where the matching is non-discrete and a sample-refinement process is proposed. The framework is demonstrated effectively on the localization of a person in an image, using a kinematic model with template features. Different priors are used on the model parameters and the results demonstrate nontrivial variations in the optimal feature hierarchy.

1 Introduction

Spatial registration is a topic which concerns many areas of computer vision including image mosaicing, structure-from-motion, medical imaging, tracking and object localization. One of the most difficult and interesting problems is that of registering a kinematic structure to a person in an image. This has many difficult aspects, such as choice of features and handling self-occlusions. While these difficulties continue to exist, the problem which we will address in this paper is that of maximizing search efficiency in registering a high dof model with known features (eg. predefined appearance templates), but without prior knowledge of the model state (see figure 1). Scenarios which benefit directly from a solution to this problem include bootstrapping a person tracker for individuals recorded in a image

database, or the re-initialization of trackers when tracking failure has been detected. A naive approach would be to search the entire kinematic state-space, which is however computationally intractable.

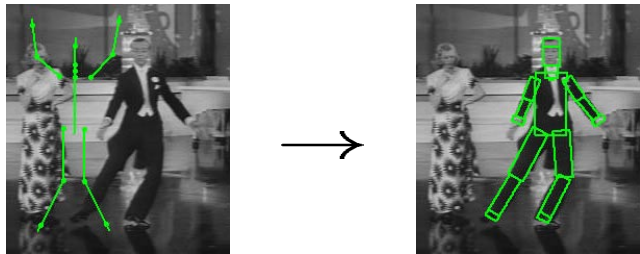


Figure 1. Registering a high dof kinematic model with known features, but without prior knowledge of the model state. Left image shows the initial state of the model, right image shows the desired state.

In this paper we will describe an approach to minimizing the amount of search in order to provide a tractable method of registration. Furthermore the analysis we provide is not limited to articulated structures but applies to *any form of spatial registration involving a model with multiple features*. Previous work on person localization is also summarized in section 7.

1.1 Minimizing Search

In situations when the prior knowledge is weak, registration almost always involves search as there will often be multiple candidates for the correct feature location. An important goal in the design of the algorithm is to minimize the amount of search required in order to maximize the efficiency of the registration process.

An existing approach to minimizing search would be to carry out the feature-matching in a *sequential* manner. The

reason for this is to reduce the uncertainty in the states of subsequent features such that the amount of search required for matching these features is reduced [20, 18].

A probabilistic estimation framework for incrementally improving the model state and covariance estimates by incorporating features sequentially was proposed by Hel-Or and Werman [11]. The framework is based on Kalman filter estimation [2] except that a prediction step is not used. It treats each feature as a separate observation which updates the model state and covariance matrix using the standard Kalman update step. This is also the framework which we will adopt in our paper, and the relevant equations are given later in (5)-(7).

Although we adopt the same estimation framework as [11], we disagree with their search strategy of selecting features sequentially along the articulated chain. Although this is an intuitive idea, it is also a *sub-optimal* strategy – the optimal strategy which we will propose in this paper can choose features on an articulated structure *out of sequential order*. Figure 4 illustrates this quite clearly.

A number of methods involve selecting features in no specific order. The various RANSAC-derived methods [8, 17] select random minimal sets of feature pairs to compute an initial estimate for the model parameters, which are then validated or invalidated based on the number of subsequent pairings admitted by this estimate. These additional feature pairs are used to improve the previous model estimate.

Other techniques offer a predefined feature ordering. Methods which adopt a multi-resolution approach [5, 13, 19, 14, 12] order features according to their resolution level – an initial model estimate is obtained at a lower resolution before proceeding to higher resolutions in order to maximize search efficiency. Similarly, this is also the case for methods which select features according to a known hierarchical decomposition of components [10].

While some of these methods provide an intuitively search-efficient feature ordering, they implicitly assume some weak generic prior for the model parameters. For example, methods which are based on always registering coarse features first do not optimally handle cases when the positions some fine-scale features are accurately known in advanced.

In the following sections we will formalize a framework for optimal feature ordering and show that it evolves dynamically according to the estimated model state and covariance. Particularly in cases where the prior encodes specific spatial information, the optimal feature ordering differs significantly from the predefined intuitive ordering.

2 Spatial Registration Framework

In this paper, we express the general spatial registration framework as follows: We start with a set of known

‘source’ features \mathcal{F} and a transformation model \mathcal{M} which maps these features into an image. Then given a target image, the goal is to match these features to their correct locations in the image and also to recover the parameters of \mathcal{M} denoted as a vector x . These features can either be prior knowledge as part of the model specification, or in the case of registering two images they represent extracted features.

Feature-based registration may be classified into two categories:

- *Feature-to-feature matching*. In this case a separate set of features is extracted from the target image. Matching is done in a discrete manner by attempting to match ‘source’ features to ‘target’ features. The features applicable for this form of matching are discrete features such as corners, edges and contours.
- *Feature-to-image matching*. Here the source features are projected into the image and compared directly. For example, template features can be matched to the image by minimizing a measure of pixel difference.

The amount of search required in the registration process depends significantly on the apriori knowledge of the model parameters. For example if x has a small prior covariance, such as in video-based tracking applications, discrete feature-matching may simply involve mapping the source features into the image and searching for the nearest target features. The model parameters may then be computed directly from these correspondences. Similarly if template features are used instead, registration may be carried out in the model state-space by locally minimizing the pixel residual error. Registration in these problems which have strong priors do not have significant search complexities and all features can be matched simultaneously.

In the case of registering a kinematic model of the figure to an image, F may be the set of template features associated with the links in the model, and M is parameterized by a vector of joint angles and link lengths. These features are not necessarily limited to a single class, as F can simultaneously include templates, corners and edges. It can also include features from different levels of resolution.

3 Analysis of Spatial Features

A feature $f \in \mathcal{F}$ is formally described by a number of attributes:

1. A function $G : x \mapsto u$ which maps the model state x to a feature state u in a common feature space.
2. A property vector ρ which allows a feature to be compared with another through a comparison function, or compared to the image.

3. Additionally for image-based features such as templates, we specify the dimensions for the **basin of attraction** in feature space. This specifies the maximum displacement between the true and predicted locations of the feature in feature space for which local optimization of the estimated location (via the maximization of a comparison function) will guarantee to converge on the true location.

In the case of discrete feature-matching, a feature comparison function $C_{ff}(\rho_i, \rho_j)$ generates a similarity measure for comparing feature pairs. In the case of feature to image matching, the comparison function $C_{fi}(\rho_j, \mathbf{u}_i, \mathbf{I})$ measures the compatibility between the feature in its current feature state with the image \mathbf{I} – it is through the maximization of this function by which the image-based features can be optimally localized.

In this paper, we assume that the correct feature pair or feature state maximizes the relevant comparison functions, ie. once all candidate features or states are tested, the correct solution will be obtained. Obviously this is not necessarily true in cases where the comparison functions generate noisy measures, and a framework for obtaining multiple-hypothesis solutions to the registration problem will be proposed in a future paper.

3.1 Matching Ambiguity of a Feature

Given the estimated model state μ and covariance Σ , we define the matching ambiguity of a feature as follows:

Definition 1 (Matching Ambiguity)

The matching ambiguity of feature f_i , denoted by α_i , is defined as the number of search operations required to find the true match with some specified minimum probability.

The idea proposed here applies the *validation gate* [3] used in extended Kalman filters. Linearizing the mapping $G_i(x)$ about μ , the covariance S_i in feature space is expressed as

$$S_i = J_i \Sigma J_i^T \quad (1)$$

where

$$J_i = \nabla G_i|_{x=\mu} \quad (2)$$

is the Jacobian. The validation gate is then the volume bounded by an equiprobability surface which may be specified as a factor ψ of standard deviations. In our experiments, the validation gates used span 2.5 standard deviations ($\psi = 2.5$).

For feature-to-feature matching, the matching ambiguity is then the number of target features which lie within the validation gate. This may be obtained by evaluating the Mahalanobis distances to potential target features and counting. Unfortunately, this is a potentially intensive computation because it would involve pairwise comparisons of features. A reasonable approximation which can be used when

target features are approximately uniformly distributed is that the matching ambiguity is proportional to the size of the validation gate, ie.

$$\alpha_i \propto (\|S_i\|)^{\frac{1}{2}} \quad (3)$$

Since in the algorithm proposed later the matching ambiguities are used to sort the features, the exact values of the matching ambiguities need not be evaluated as long as they can be ranked in the right order.

For feature-to-image matching, the matching ambiguity is the number of minimally-overlapping regions which have the same dimensions as the basin of attraction that would fit into the validation gate. This can be approximately computed through the following steps:

1. obtain the eigenvalues e_j and eigenvectors v_j to the covariance matrix S_i ;
2. calculate the span of the basin of attraction b_j along each of the v_j directions;
3. the matching ambiguity is then computed as

$$\alpha_i \approx \prod_j \text{ceil} \left(\psi \frac{\sqrt{e_j}}{b_j} \right) \quad (4)$$

where $\text{ceil}(\cdot)$ rounds fractional values up to the next integer.

Figure 2 illustrates the concept of matching ambiguity for the two separate cases.

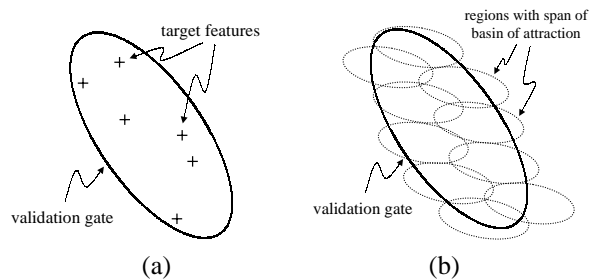


Figure 2. Matching ambiguity. Feature-to-feature matching: (a) shows the target features located within the validation gate of a source feature; the matching ambiguity in this case is the number of candidates to be search, which is 6. Feature-to-image matching: (b) shows minimally overlapping regions with the span of the basin of attraction covering the validation gate; the matching ambiguity here is the number of regions required, which is 10.

4 Sequential Registration with Dynamic Feature Ordering

As each feature is used in the estimation step during sequential search, the model state becomes increasingly more accurate and the state covariance decreases. This observation can be formalized using the standard Kalman filter covariance update step (6). Correspondingly based on (1), the size of the validation gates for each feature decreases, thereby leading to a reduction in the matching ambiguities α_i .

Assuming that the intention is to use all available features sequentially in the registration process¹, the total number of search operations involved can be minimized by selecting features and estimating the model parameters in the algorithm described below, which we call the *Coupled Dynamic Feature Ordering and Registration (2DYFOR)* algorithm.

The 2DYFOR Algorithm

1. Set the list of used features L as empty.
2. Compute the matching ambiguities α_i for all unused features.
3. Select the feature f_b for which α_b is the smallest matching ambiguity.
4. Carry out the necessary α_b search operations to recover the optimal feature state \mathbf{u}_b . This is the minimum number of search operations which have to be performed to register a feature.
5. The optimal feature state \mathbf{u}_b and the associated observation covariance S_b is used to improve the model state and covariance by applying the standard Kalman filter update steps:

$$\boldsymbol{\mu}_k = \boldsymbol{\mu}_{k-1} + \mathbf{K}_k(\mathbf{u}_b - \mathbf{J}_k \boldsymbol{\mu}_{k-1}) \quad (5)$$

$$\boldsymbol{\Sigma}_k = \boldsymbol{\Sigma}_{k-1} - \mathbf{K}_k \mathbf{J}_k \boldsymbol{\Sigma}_{k-1} \quad (6)$$

where the subscript k denote the sequential update index and \mathbf{K}_k is the Kalman gain given by

$$\mathbf{K}_k = \boldsymbol{\Sigma}_{k-1} \mathbf{J}^T (\mathbf{J} \boldsymbol{\Sigma}_{k-1} \mathbf{J}^T + S_b)^{-1} \quad (7)$$

6. Append f_b to the L .
7. If all features have been used, stop; otherwise return to step 2.

¹This is not always true, as there may be formulations based on sampling of features [1] or *optimal stopping* strategies [7].

At the end of the registration process, the feature list L contains the *feature hierarchy*. The feature hierarchy represents the optimal sequential ordering of features and is dependent on the prior model state and covariance, as well as the accuracy of registering each feature. The feature hierarchy has to be formed dynamically as part of the estimation process. While the predefined feature orderings used in the algorithms may be reasonably efficient in typical situations, the optimal feature hierarchy can often be found at negligible cost using the 2DYFOR algorithm. Furthermore, the dynamic feature ordering copes even when the prior knowledge changes significantly – using the original predefined feature ordering may not take full advantage of the additional prior knowledge for increasing search efficiency.

5 Search Method for Feature-to-Image Matching

While it may be straightforward to carry out feature-to-feature matching based on the 2DYFOR algorithm described in section 4, implementing feature-to-image matching is more complex and will be discussed in further detail in the following section.

5.1 The Sample-Refinement Approach to Search

In feature-to-image matching problems an attempt to recover the optimal feature can be made by locally maximizing a similarity measure obtained from the comparison function C_{FI} , which is described by the term *refinement*. However the starting feature state must be within the basin of attraction of the correct solution for the refinement process to succeed. Hence it is necessary to generate a number of starting points (termed as *samples*) for the refinement step, spaced at intervals corresponding to the span of the basin of attraction, in order to guarantee that the optimal feature state will be found.

One issue which arises for this method is whether samples should be obtained as feature state-vectors or model state-vectors. It turns out that generating the samples in model state-space has two advantages:

- If the non-linearity between the feature and model state-spaces is significant with respect to the validation gate, and if the feature state-space is not mapped entirely by the model state-space, sampling in model state-space ensures the starting feature states are valid for the model, even if the spacing of the samples is inaccurate due to the nonlinearity.
- Often the refinement step can be improved by using previously registered features as well, especially

when the comparison function generates noisy measures. Using the additional features improves the robustness to the noise.

Using notation from section 3.1, the desired displacement between samples in feature space along the eigenvector direction \mathbf{v}_j is $b_k \mathbf{v}_j$. The equivalent displacement in model state-space may be found by the following analysis:

$$\begin{aligned} \mathbf{S} \mathbf{v}_j &= e_k \mathbf{v}_j \\ \Leftrightarrow \mathbf{J} \Sigma \mathbf{J}^T \mathbf{v}_j &= e_k \mathbf{v}_j \\ \Leftrightarrow \mathbf{J} \begin{pmatrix} b_k \\ e_k \end{pmatrix} \Sigma \mathbf{J}^T \mathbf{v}_j &= b_k \mathbf{v}_j \end{aligned} \quad (8)$$

Noting that \mathbf{J} represents the linearized mapping between model and feature state-spaces, it may be observed that using the term in the parentheses on the LH side of (8) as the displacement in model state-space generates the desired displacement in feature space, given by the RH side of (8).

The samples in the model state-space are then generated by integral vector sum combinations of the displacement vectors associated with each eigenvector direction in feature space, such that the span of the validation gate in feature space is covered by these samples. Once the initial samples have been generated in the model state-space, each sample is then refined through the local maximization process. The optimal solution is then taken to be the refined sample which generates the maximum similarity measure from the comparison function.

6 Kinematic Model Registration

The model used in our experiments is the 2D Scaled-Prismatic Model (SPM) proposed by Morris and Rehg [15]. The kinematic model lies in the image plane, with each link having a degree-of-freedom (dof) in rotation and another dof in length. The model is parameterized by a state-space which encodes a global 2D translation, joint angles and link lengths. Each link is associated with a template which describes the appearance of the link. The approach used to locally refine the state of the kinematic model is to minimize the SSD error between the templates and the image using the Gauss-Newton method. The model used for the Fred Astaire image has 19 states and 16 template features, while the model used for the walking figure has 8 states (arms are ignored and link lengths are fixed) and 10 template features.

In our experiments which involve localizing the figure with minimal prior knowledge, the model state is initialized as denoted by the pose of the stick figure in the leftmost images of figure 3. The prior covariance is set as a diagonal matrix with standard deviations of 50 pixels for global x

translation, 20 pixels for global y translation, 2 radians for joint angles and 10 pixels for link lengths. The only strong prior is that the torso is approximately upright as we wish to restrict our search to upright figures. For each template, the basin of attraction for the refinement step is set to be its minimum dimension at present, although a more formal analysis may be applied in the future based on the spatial frequency content of the templates.

The sequences shown from left to right in figure 3 illustrate the feature ordering which arises in the registration process. The feature ordering obtained in these instances is similar to a size-based ordering, except in our algorithm the ordering is done both automatically and dynamically. The registration localizes the figure well despite the high dimensionality of the figure model and the weak prior knowledge.

In figure 4, we show the results obtained when various forms of strong prior knowledge are available which is captured in the prior covariance. For the top test image, the feet template positions are accurately positioned, while for the bottom test image the front shin template position is accurately positioned. A 5-pixel standard deviation is used for these constraints. The feature ordering in these instances differs significantly from the ordering obtained in figure 3. Also notice from the third-from-left image in the top sequence, the optimal feature ordering does not propagate along the articulated chains which contradicts the proposed heuristic of Hel-Or and Werman [11].

The top test images generally took approximately one to two minutes for the localization, while the bottom test images took approximately twenty seconds because of the simpler model used. In both instances the number of samples used for initializing the search of individual templates appears to be significantly more than is necessary, which is due to the conservative estimates for the span of the basins of attraction in the refinement process. Hence there is still significant room to improve the efficiency of the registration.

7 Previous Work on Person Localization

While current systems are good at detecting humans as moving blobs, few solutions are available when the *full articulated pose* of a person including the position of the arms and legs is required. For example, the full-body tracking systems described in [6, 4] currently depend on manual initialization of a figure model in the first image frame. The method in [9] has automatic initialization but requires an accurate background model for segmentation and restrictive assumptions on limb positions. An interesting method for registering articulated structures based on EM motion segmentation is proposed in [16], but has only been applied to low dimensional models. Furthermore, many systems depend on independent motion of the figure for segmentation

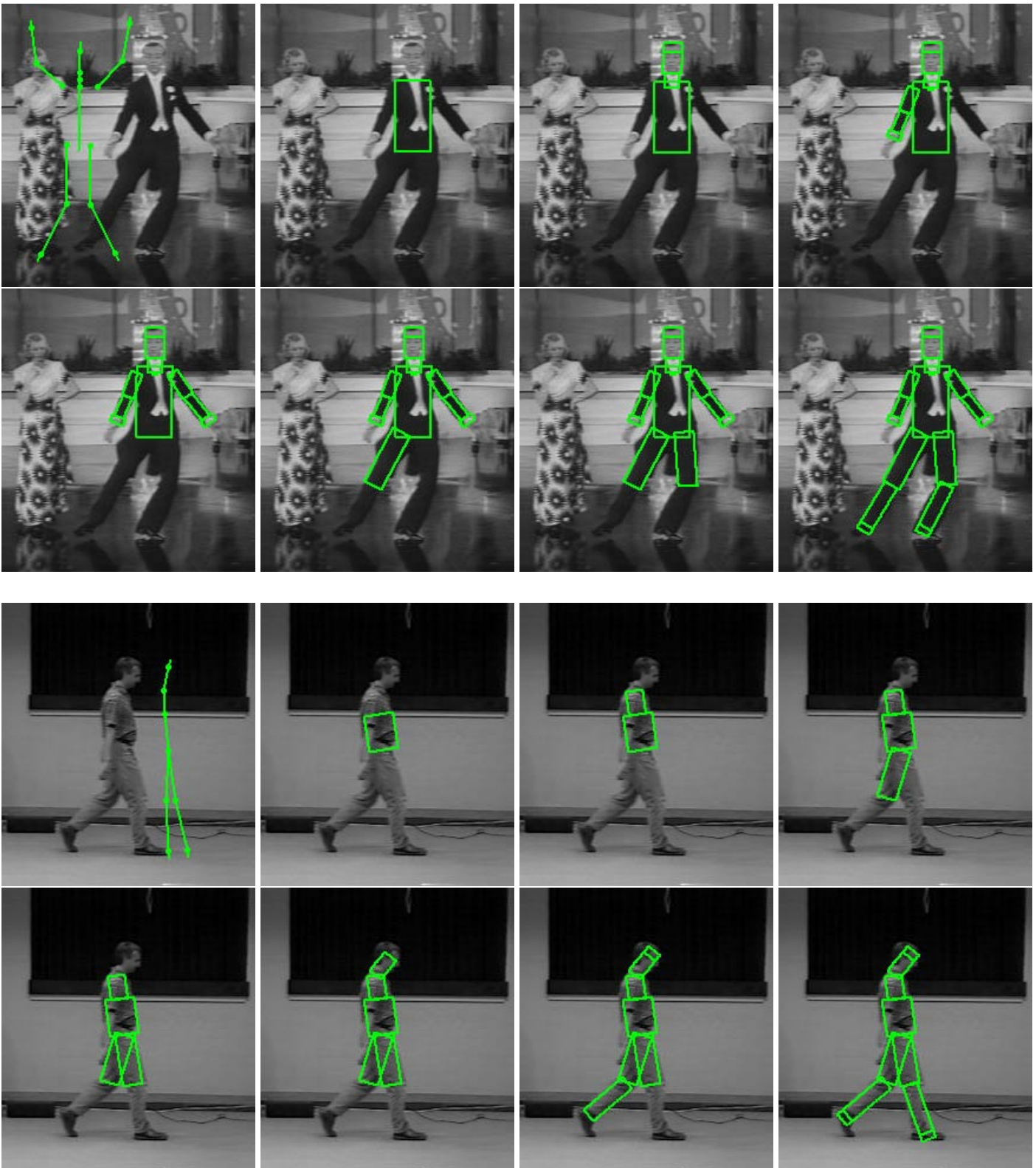


Figure 3. Results obtained using weak generic priors. The top-left image of each block show the prior state of the kinematic model which is represented by a stick figure. Weak generic priors are used in these two test cases. As the coupled dynamic feature ordering algorithm is iterated, the next template feature selected is the one requiring the least amount of search operations for registration. The left-to-right sequences show the feature ordering.

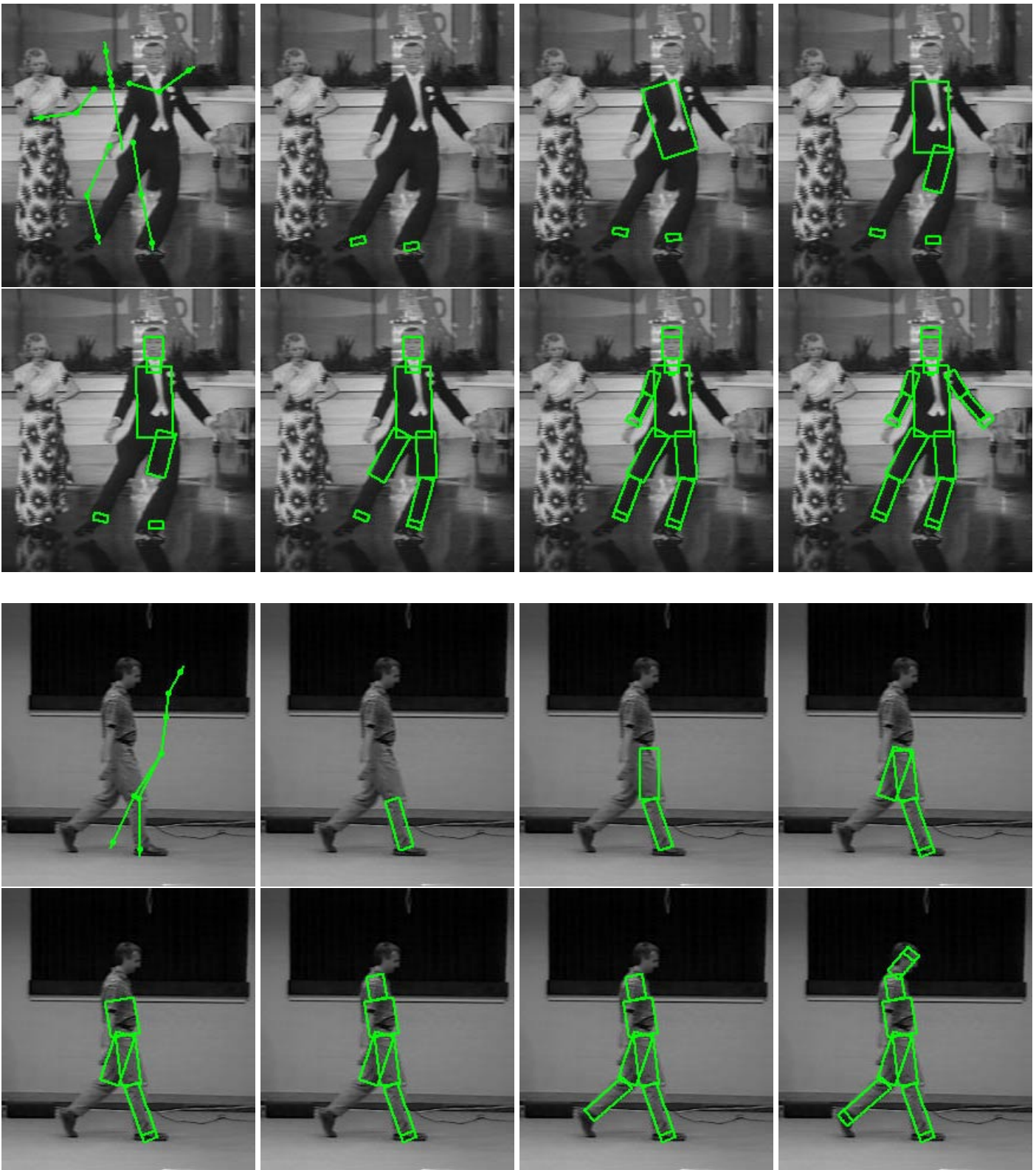


Figure 4. Results obtained using some strong prior knowledge. In the top test case, the position of the two feet templates as assumed to be known accurately. In the bottom test case, the position of the front shin template is known accurately. The feature ordering obtained when strong prior knowledge is available can be significantly different from the ordering with weak generic priors. Note the registration of the left leg in the top test case is *corrected* as more features are integrated into the estimation.

and registration (eg. those based on optic-flow), and cannot be applied to static images or to still figures.

8 Conclusions and Future Work

This paper presents a general algorithm for maximizing search efficiency for a sequential registration problem. At each iteration in the registration process, the feature with the minimum matching ambiguity is selected and used in the estimation. The means of computing the matching ambiguity is discussed separately for problems involving feature-to-feature matching and feature-to-image matching. Additionally for the latter cases, a search method is proposed which is based on generating samples in the model state-space such that at least one sample will fall into the basin of attraction for the correct solution. These samples are further refined to obtain the optimal feature state which is then used to improve the model state estimation.

The current application of this framework is registering kinematic models to human figures in images. Despite the high dimensional kinematic models used and the weak priors assumed, the framework is able to register the models accurately and efficiently. Furthermore, the optimal feature ordering is shown to be significantly different when additional strong prior knowledge can be used to constraint the searches.

There are a number of situations when the strong prior knowledge is available. For example, partial figure tracking failure may result from the occlusion of the torso by a shoulder-height object, although the person's head will still be well tracked; in this instance the smaller head feature will provide strong constraints on the location of the torso. Another example would be a semi-automated tracking system (eg. for video editing) where the user provides partial registration of a model in a video sequence. In both scenarios, our proposed method would efficiently obtain the remaining correspondences by utilizing the available prior knowledge.

Currently, our system is based on the assumption that the feature searches always return the correct registration. However, multiple registration candidates may arise as a result of clutter and self-occlusion. A future research direction would be to combine our registration approach with the multiple-hypothesis probabilistic framework proposed in [6] to cope with these problems.

References

- [1] M. Akra, L. Bazzi, and S. Mitter. Sampling of images for efficient model-based vision. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 21(1):4–11, 1999.
- [2] B. D. Anderson and J. B. Moore. *Optimal Filtering*. Prentice-Hall, 1979.
- [3] Y. Bar-Shalom and T. E. Fortmann. *Tracking and Data Association*. Academic Press, 1988.
- [4] C. Bregler and J. Malik. Estimating and tracking kinematic chains. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 8–15, Santa Barbara, CA, 1998.
- [5] T.-J. Cham and R. Cipolla. A statistical framework for long-range feature matching in uncalibrated image mosaicing. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 442–447, Santa Barbara, CA, 1998.
- [6] T.-J. Cham and J. Rehg. A multiple hypothesis approach to figure tracking. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, volume II, pages 239–245, Fort Collins, Colorado, 1999.
- [7] M. H. DeGroot. *Optimal Statistical Decisions*. McGraw-Hill, 1970.
- [8] M. Fischler and R. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Comm. ACM*, 24(6):381–395, June 1981.
- [9] I. Haritaoglu, D. Harwood, and L. Davis. W⁴: Who? When? Where? What? A real time system for detecting and tracking people. In *Proc. Intl. Conf. on Automatic Face and Gesture Recognition*, pages 222–227, Nara, Japan, 1998.
- [10] A. Hauck, S. Lanser, and C. Zierl. Hierarchical recognition of articulated objects from single perspective views. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 870–876, San Juan, Puerto Rico, 1997.
- [11] Y. Hel-Or and M. Werman. Constraint fusion for recognition and localization of articulated objects. *Int. Journal of Computer Vision*, 19(1):5–28, 1996.
- [12] M. Luetzgen, W. Karl, A. Willsky, and R. Tenney. Multiscale representations of markov random fields. *IEEE Trans. on Signal Processing*, 41(12):3377–3396, 1993.
- [13] S. Marapane and M. Trivedi. Multi-primitive hierarchical (MPH) stereo analysis. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 16(3):227–240, 1994.
- [14] E. Mémin and P. Pérez. A multigrid approach for hierarchical motion estimation. In *Proc. Intl. Conf. on Computer Vision*, pages 933–938, Bombay, India, 1998.
- [15] D. Morris and J. Rehg. Singularity analysis for articulated object tracking. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 289–296, Santa Barbara, CA, 1998.
- [16] H. A. Rowley and J. M. Rehg. Analyzing articulated motion using expectation-maximization. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 935–941, San Juan, Puerto Rico, 1997.
- [17] P. Torr. *Motion Segmentation and Outlier Detection*. PhD thesis, University of Oxford, 1995.
- [18] K. Toyama and G. Hager. Incremental focus of attention for robust visual tracking. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 189–195, San Francisco, CA, 1996.
- [19] J. Weng, N. Ahuja, and T. Huang. Matching two perspective views. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 14(8):806–825, 1992.
- [20] L. Wixson. *Gaze Selection for Visual Search*. PhD thesis, Department of Computer Science, University of Rochester, 1994.