Optimization of Symmetric Transfer Error for Sub-frame Video Synchronization

Meghna Singh[1], Irene Cheng[2], Mrinal Mandal[1], and Anup Basu[2]

[1] Department of Electrical and Computer Engineering
[2] Department of Computing Science
University of Alberta, Edmonton, Alberta, Canada

Abstract. In this work we present a method to synchronize video sequences of events that are acquired via uncalibrated cameras at unknown and dynamically varying temporal offsets. Unlike existing methods that synchronize videos of similar events (i.e., videos related to each other through the motion in the scene) up to an integer alignment, we establish sub-frame video synchronization. While contemporary synchronization algorithms implement a unidirectional alignment which biases the results towards a single reference sequence, we adopt a bi-directional or symmetrical alignment approach that results in a more optimal synchronization. To this end, we propose a novel symmetric transfer error which is dynamically minimized, and reduces the propagation of error from feature extraction and spatial mapping into temporal synchronization. The advantages of our approach are validated by tests conducted on (publicly available) real and synthetic sequences. We present qualitative and quantitative comparisons with another state-of-the-art algorithm. A unique application of this work in generating high-resolution 4D MRI data from multiple low-resolution MRI scans is described.

1 Introduction

Synchronization of video sequences plays a crucial role in applications such as super-resolution imaging [1][2], 3D visualization [3], robust multi-view surveillance [4] and mosaicking [5]. Most video synchronization algorithms deal with video sequences of the same scene and hence assume that the temporal offset between the video sequences does not change over time, i.e., a simple temporal translation is assumed. Video synchronization, however, is not limited to aligning sequences of the same scene, and can be extended to find the spatio-temporal alignment between related scenes, for applications such as video search, video comparison and enhanced video generation. In such scenarios, the temporal offset between the video sequences is dynamically changing, and cannot be estimated by a translational offset. The closest related works in synchronization of video sequences that are related by dynamically varying temporal offsets are limited to integer frame alignment of video sequences [6][7].

Another drawback identified with contemporary works is the asymmetric or unidirectional nature of the alignment, due to which error or uncertainty in

D. Forsyth, P. Torr, and A. Zisserman (Eds.): ECCV 2008, Part II, LNCS 5303, pp. 554–567, 2008.

feature extraction in the reference sequence propagates into the synchronization result. Our contribution lies in formulating the synchronization problem as the iterative minimization of a *symmetric transfer error* (STE), which allows us to compute *sub-frame accurate* synchronization of video sequences that have a *dynamically varying temporal offset* between them. In addition, the method of minimizing STE allows us to reduce the occurrence of singularities[1] in the synchronization.

The rest of this paper is organized as follows. Section 2 outlines related work in video synchronization, and justifies the necessity for this work. In Section 3, we present the symmetric transfer error and an optimization strategy to minimize it. In Section 4, we discuss our experimental setup and show comparative results with a rank-constraint based approach. Application of our work in 4D MRI visualization is presented in Section 5. Finally, summary and conclusions are given in Section 6.

2 Review of Prior Research

Past literature in video synchronization and temporal registration can be broadly classified into two categories — video sequences of the same scene or video sequences of similar scenes; differing primarily on the assumptions made with respect to the temporal offset between sequences. It can be seen from Table 1, that our work addresses all three scenarios (view-invariance, dynamic time shifts and sub-frame accuracy), while previous works have only addressed a subset of these scenarios.

2.1 Video Synchronization of Same Scene

In synchronizing videos of the same scene, the temporal offset is considered to be an affine transform [1], such as $t' = s.t + \Delta t$, where s is the ratio of frame rates and Δt is a fixed translational offset. Dai et al. [9] use 3D phase correlation between video sequences, whereas Tuytelaars et al. [10] compute synchronization by checking the rigidity of a set of five (or more) points. Tresadern et al. [11] also follow a similar approach of computing a rank-constraint based rigidity measure between four non-rigidly moving feature points. Lee et al. [4] derive the spatial relationship between multiple surveillance cameras viewing the same (or partially overlapping) scene by matching motion trajectories captured in the video sequences to planar models. Caspi et al. [1] recover the spatial and temporal relation between two sequences by minimizing the SSD error over extracted trajectories that are visible in both of the sequences. Carceroni et al. [12] extend [1] to align sequences based on scene points that need to be visible only in two

[1] Singularities [8] occur when multiple frames in the target sequence map to the same frame in the reference sequence. Such a situation can occur when the temporal speed of an event in one sequence is significantly slower than another sequence. However, in most cases this slowing down of the event should lead to a sub-frame mapping and not singularities.

Table 1. Review of related work. Legend: D.S.-Dynamic time shift, V.I.-View Invariant, S.A.-Sub-frame Accuracy.

Author	Scene	D.S.	V.I	S.A.	Author	Scene	D.S.	V.I	S.A.
Caspi et al. [2] [1]	Same	X	$\surd$	$\surd$	Tresadern et al. [11]	Same	X	$\surd$	$\surd$
Lei et al. [13]	Same	X	$\surd$	X	Carceroni et al. [12]	Same	X	$\surd$	$\surd$
Wolf et al. [14]	Same	X	$\surd$	X	Pooley et al. [15]	Same	X	$\surd$	$\surd$
Dai et al. [9]	Same	X	$\surd$	$\surd$	Tuytelaars et al. [10]	Same	X	$\surd$	X
Singh et al. [3]	Same	X	X	$\surd$	Lee et al. [4]	Same	X	$\surd$	X
Perperidis et al. [16]	Diff	$\surd$	X	X	Rao et al. [6]	Diff	$\surd$	$\surd$	X
Proposed	Diff	$\surd$	$\surd$	$\surd$	Giese et al. [17]	Diff	$\surd$	X	X

consecutive frames. Singh et al. [3] also build on [1] to develop parameterizable event models of discrete trajectories and align the event models based on an SSD measure for sub-frame synchronization. However, they do not address the problem of view-invariance and dynamic temporal offsets.

2.2 Video Synchronization of Different Scenes

When aligning video sequences of different scenes, albeit sequences correlated via motion, one has to factor in the dynamic temporal scale of activities in the video sequences. Giese and Poggio [17] approach alignment of activities of different people by computing a dynamic time warp between the feature trajectories. They do not address the problem when the activity sequences are from varying viewpoints and their approach is a one-to-one frame correspondence. Perperidis et al. [16] attempt to locally warp cardiac MRI sequences, by extending Caspi's work [1] to incorporate spline based local alignment. Though their approach does lead to good alignment of time-varying sequences, it has two main drawbacks: (i) the computation space for spline based registration is quite large and the authors need to compute points of inflexion in the cardiac volume change; and (ii) the alignment is still a one-to-one frame correspondence and not sub-frame accurate. Others, such as Rao et al. [6] use rank constraint as the distance measure in a dynamic time warping algorithm to align multiple sequences; this is the first work that can deal with video sequences of correlated activities, hence we will present comparisons with their rank constraint based (RCB) algorithm.

3 Proposed Approach

In this work we assume that the frame rates of the cameras are identical and fixed throughout the acquisition, the scene is planar and the backgrounds in both the scenes have sufficient static points that can be extracted using view-invariant feature detectors [18] to estimate the spatial relationship between the two scenes. These assumptions are not restrictive since the synchronization algorithm can be adapted to account for epipolar geometry of non-planar scenes and

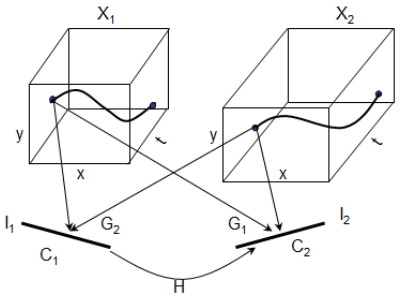


Fig. 1. Illustration of two distinct scenes acquired using two distinct cameras. The projections (ghosts) of scenes onto the reciprocal cameras are also shown.

robust correspondence algorithms can be used for wide-baseline cameras [19]. Additionally, we assume that a single feature trajectory of interest is available to us[2] such that the beginning and end points of the activity are marked in the trajectory, similar to the assumption made in [6].

Given that we have two cameras C_1 and C_2, as shown in Fig.1, that view two independent scenes of similar activities; C_1 views scene X_1 and acquires video I_1, and C_2 views scene X_2 and acquires a video sequence I_2. Features, F_1 and F_2, are extracted and tracked in both the acquired video sequences. A single feature trajectory is used to illustrate this in the figure. On their own these feature trajectories are discrete representations of the event in the scene, and we need to interpolate between these discrete representations. Most algorithms use linear interpolation, however, we use the method proposed by Singh et al. [3] to generate continuous models from discrete points. The continuous feature trajectories are represented as $\mathcal{F}_1$ and $\mathcal{F}_2$. Note that we store these continuous models as discrete matrices in computer memory, albeit the temporal resolution of the continuous models is much higher than the discrete models. Hence, we index into $\mathcal{F}$ using integer numbers.

The spatial relationship (homography H) between the two scenes is computed by using the Direct Linear Transform (DLT) algorithm to initialize the non-linear least squares (Levenberg-Marquardt) computation of the homography [20]. As the feature trajectories are dynamically offset from each other (as illustrated in Fig.1), we cannot directly optimize for both the homography and temporal offset as proposed by Caspi et al. [1]. Instead, we project the trajectory from Scene 1 to Scene 2, as if Camera C_2 did view Scene 1 (and vice versa). Note that the accuracy of this projection is subject to noise in the extracted trajectories as well as error in the homography computation. We call these projections the *ghosts* of the trajectories, represented as $\mathcal{G}$.

[2] In general videos, multiple object trajectories will be generated and an additional task of the synchronization algorithm will be to find corresponding feature trajectories in the multiple video sequences. This is an open problem in vision research and one that we will not address in this paper.

$$\mathcal{G}_1 = H_{1\to 2} \cdot \mathcal{F}_1$$
$$\mathcal{G}_2 = H_{2\to 1} \cdot \mathcal{F}_2 \tag{1}$$

The aim of the synchronization algorithm is to temporally align a trajectory in Scene 1 with the ghost of the trajectory from Scene 2, and vice versa. Current dynamic offset synchronization algorithms, *e.g.* [6] and [17], synchronize discrete feature trajectories to a frame-by-frame correspondence by only computing a unidirectional alignment. For example, they assume Trajectory 1 to be the reference, and warp Trajectory 2 towards it. This unidirectional alignment biases the synchronization towards the reference sequence. Feature extraction and tracking errors in the reference sequence now propagate unchecked into the synchronization. We demonstrate that a more symmetric approach will not only mitigate such an error propagation, but will also result in better sequence synchronization. Next, we introduce our symmetric optimization approach.

3.1 Alignment as the Minimization of Symmetric Transfer Error

Feature trajectories and associated event models are computed independently for each sequence; introducing a variability in the sequence alignment, such that frame alignment from Sequence 1 to Sequence 2 is not identical to frame alignment from Sequence 2 to Sequence 1; i.e., if $\mathcal{W}_{1,2}$ is a mapping from $\mathcal{F}_1$ in Sequence 1 to the ghost $\mathcal{G}_2$ of Sequence 2 and $\mathcal{W}_{2,1}$ is a mapping from $\mathcal{F}_2$ in Sequence 2 to the ghost $\mathcal{G}_1$ of Sequence 1, then $\mathcal{W}_{1,2} \neq \mathcal{W}_{2,1}$. This asymmetry is illustrated with a hypothetical example in Fig.2, where $\mathcal{W}_{1,2}$ and $\mathcal{W}_{2,1}$ are two reciprocal alignments computed between the sequences. Asymmetry in mappings for real video sequences is shown in Fig.9. If indexing function $\mathcal{M}(\mathcal{W}_{1,2}, i)$ indexes the mapping $\mathcal{W}_{1,2}$ and returns the frame in Sequence 2 corresponding to the i^{th} frame in Sequence 1 (similarly $\mathcal{M}(\mathcal{W}_{2,1}, i)$ indexes mapping $\mathcal{W}_{2,1}$ and returns the corresponding frame in this mapping), then the symmetric transfer error (STE) for the i^{th} frame (where $i \in \mathbb{R}$) is defined as follows:

$$\mathcal{E}(i) = |\mathcal{M}(\mathcal{W}_{1,2}, i) - \mathcal{M}(\mathcal{W}_{2,1}, i)| \tag{2}$$

Let us illustrate (2) with an example. Suppose in Fig.2, for the i^{th} frame in Sequence 1, mapping $\mathcal{W}_{1,2}$ reports the corresponding frame in Sequence 2

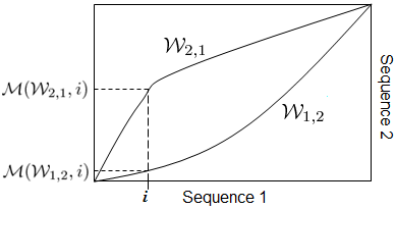


Fig. 2. Computing symmetric transfer error for a single frame 'i' in Sequence 1

to be the 7^{th} frame, and mapping $\mathcal{W}_{2,1}$ reports the corresponding frame to be the 10^{th} frame. Then, STE for the i^{th} frame is $\mathcal{E}(i) = 3$. STE for the entire sequence is then defined as the summation in (3), where N is the length of $\mathcal{F}_1$ and $\mathcal{G}_1$ of Sequence 1 and M is the length of $\mathcal{F}_2$ and $\mathcal{G}_2$ of Sequence 2. Intuitively, minimizing STE (3) for synchronization is akin to computing an optimal compromise between the reciprocal mappings $\mathcal{W}_{1,2}$ and $\mathcal{W}_{2,1}$.

$$\mathcal{E} = \sum_{i=1..max(N,M)} |\mathcal{M}(\mathcal{W}_{1,2}, i) - \mathcal{M}(\mathcal{W}_{2,1}, i)| \ . \tag{3}$$

The mapping functions $\mathcal{W}_{1,2}$ and $\mathcal{W}_{2,1}$ are computed using a regularized dynamic time warping (DTW) approach. The implementation of DTW via dynamic programming, factors in boundary conditions, continuity and monotonicity of the mapping function. An important distinction to be made here is that the STE is not the same as a symmetric implementation of DTW (details on this distinction are provided online at [21]). While (4)-(7) detail how $\mathcal{W}_{1,2}$ is computed; the same apply to $\mathcal{W}_{2,1}$ with suitable substitutions made for $\mathcal{F}_2$ and $\mathcal{G}_1$. We build a cost matrix $\mathcal{D}(n, m)$ $\forall n \in [1..N]$ and $\forall m \in [1..M]$, where N is the length of event model $\mathcal{F}_1$ for Sequence 1 and M is the length of event model $\mathcal{F}_2$ for Sequence 2, as follows:

$$\mathcal{D}(n, m) = \|(\mathcal{F}_1(n) - \mathcal{G}_2(m))\|^2 + w(\|\partial\mathcal{F}_1(n) - \partial\mathcal{G}_2(m)\|^2) \ , \tag{4}$$

$$\partial\mathcal{F}(k) = \frac{\mathcal{F}(k) - \mathcal{F}(k-1) + [\mathcal{F}(k+1) - \mathcal{F}(k-1)]/2}{2} \ . \tag{5}$$

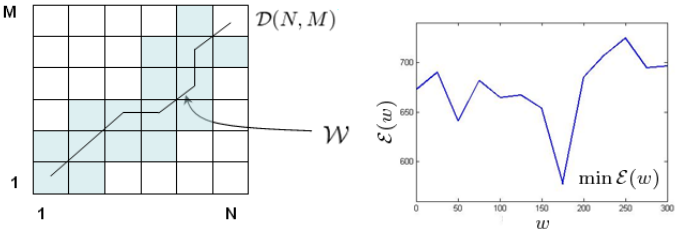
In (4), w is the weight assigned to the regularization function. The motivation behind regularization of the cost function is two fold: (i) it allows us to factor in a smoothness constraint on the warping and (ii) it also reduces the occurrence of singularities in the mapping. The mapping function $\mathcal{W}$, is computed by traversing the path of minimum cost in the cost matrix $\mathcal{D}$ (illustrated in Fig.3(a)), with n and m initialized to N and M respectively, as follows:

$$\mathcal{W}_{1,2}(n, m) = \mathcal{D}(n, m) + \min(\phi) \ , \tag{6}$$

$$\phi = [\mathcal{W}_{1,2}(n-1, m), \mathcal{W}_{1,2}(n-1, m-1), \mathcal{W}_{1,2}(n, m-1)] \ . \tag{7}$$

In (7) we consider a neighborhood of three frames (similar to [6]), however, this neighborhood can be extended. It can be seen from (3)-(6) that STE is dependent on the regularization weight w, and minimization of the STE with respect to w (8) optimizes the sequence synchronization. An example of STE values computed for varying values of w for two synthetic trajectories are shown in Fig.3(b).

$$\mathcal{W}_{\text{opt}} = arg\min_{w} \sum_{i=1..max(N,M)} |\mathcal{M}(\mathcal{W}_{1,2}, i) - \mathcal{M}(\mathcal{W}_{2,1}, i)| \tag{8}$$



(a) (b)

Fig. 3. (a) Illustration of computing the mapping function $\mathcal{W}$ from the cost matrix $\mathcal{D}$. (b) Plot of STE $\mathcal{E}(w)$ versus regularization weight w for two synthetic sequences of length 100 and 140 frames. $min(\mathcal{E})$ is also indicated.

Equation (8) enforces the minimization of the difference between the two mappings $\mathcal{W}_{1,2}$ and $\mathcal{W}_{2,1}$ in Fig.2. The advantage of optimizing a symmetric measure, as opposed to an asymmetric measure, is validated experimentally on real video sequences. The pseudocode of the proposed algorithm is given in Table 2. For our experiments the variables in the pseudocode min_w, max_w, and $iter\_step$, were empirically determined and set to 0, 300 and 25 respectively.

4 Experiments and Comparative Analysis

We tested our method on both synthetic and real image sequences. We also implemented the rank-constraint based (RCB) algorithm as described in [6] that deals with aligning videos of similar events. While the RCB method cannot compute sequence alignment to sub-frame accuracy, we still compare our method with it for integer alignment. Our test cases and results are presented next. All experiments were run on a 3.2GHz Intel Pentium IV processor with 1GB RAM using MATLAB 7.04. Excluding the preprocessing time, the STE algorithm took 1.79 seconds to run through all iterations, to compute the optimal alignment between two synthetic sequences of length 100 and 140. The processing time for real sequences of length 84 and 174 was 1.93 seconds.

4.1 Synthetic Sequences

In synthetic tests, we generate planar trajectories, 100 frames long, using a pseudo-random number generator (modified version of Marsaglia's subtract with borrow algorithm). These trajectories are then projected onto two image planes using user defined camera projection matrices. The camera matrices are designed so that the acquisition emulates a homography, and are used only for generation purpose and not thereafter. A time warp is then applied to a section of one of the trajectory projections, such that its length now becomes 140 frames. Both the RCB and the STE methods are then applied to the synthetic trajectories to

Table 2. Pseudocode of the proposed algorithm

Pre-preprocessing
Extract feature trajectories F_1 and F_2
Compute Event models $\mathcal{F}_1$ and $\mathcal{F}_2$
Compute Homography H
Project Event models to derive ghosts $\mathcal{G}_1$ and $\mathcal{G}_2$
Set max_w, min_w, iter_step

Iterate to minimize STE
For w=min_w : iter_step : max_w
Compute regularized warp $\mathcal{W}_{1,2}$
1. Compute cost metric $\mathcal{D}(n,m) = \|(\mathcal{F}_1(n) - \mathcal{G}_2(m))\|^2 + w(\|\partial\mathcal{F}_1(n) - \partial\mathcal{G}_2(m)\|^2)$
2. Compute mapping as the path of minimum cost $\mathcal{W}_{1,2}(n,m) = \mathcal{D}(n,m) + \min(\phi)$
$\phi = [\mathcal{W}_{1,2}(n-1,m), \mathcal{W}_{1,2}(n-1,m-1), \mathcal{W}_{1,2}(n,m-1)]$
Compute regularized warp $\mathcal{W}_{2,1}$
Follow step 1-2 above for $\mathcal{F}_2$, $\mathcal{G}_1$
Compute Symmetric Transfer Error $\mathcal{E} = \sum_{i=1..max(N,M)} |\mathcal{M}(\mathcal{W}_{1,2}, i) - \mathcal{M}(\mathcal{W}_{2,1}, i)|$
$\mathcal{M}$ is a simple indexing function.
End

Find $\min \mathcal{E}(w)$
Report $\mathcal{W}_{\text{opt}}$ corresponding to $\min \mathcal{E}(w)$

compute the alignment between them. This process is repeated on 100 synthetic trajectories. Fig. 4 shows some synchronization results with synthetic trajectories. Additional synchronization results are available online at [21], where it can be seen that for simple trajectories, the STE and RCB methods result in comparable synchronization, close to the correct alignment. However, as the complexity of the synthetic trajectory begins to increase, the RCB method starts producing erroneous alignments while the STE method continues to compute synchronization that closely matches the actual synchronization. On average, the STE algorithm made 34% less errors in computing the synchronization when compared to the RCB method. To test for sub-frame alignment, we divided the projected trajectories into two parts, such that F_1 contains every even trajectory point and F_2 contains every odd trajectory point. The computed synchronization between these two trajectories by the STE method accurately placed frames in Sequence F_2 at 0.5 frame intervals from F_1.

Table 3. Synchronization Errors for RCB and STE methods for noisy trajectories

Noise σ^2	STE	RCB
0.0001	281	384
0.001	326	470
0.01	400	755
0.1	623	1199

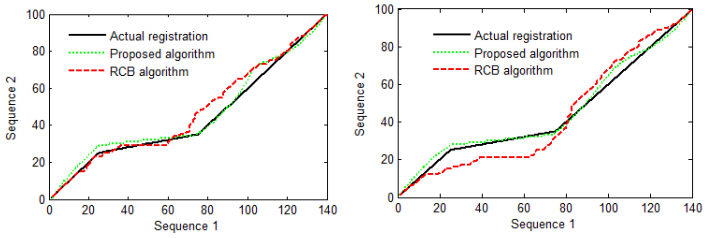


Fig. 4. (a)-(b) Results of synchronization of synthetic trajectories using proposed approach and RCB approach

We also tested the effect of noisy trajectories on our synchronization approach. Normally distributed and zero mean noise with various values of variance (σ^2) was added to the synthetic feature trajectories. The results of synchronization of noisy trajectories with both the RCB and STE approach are shown in Table 3, where the sum of absolute differences between the actual and computed frame correspondence is reported as the synchronization error. The STE approach is affected slightly by the addition of noise, however, the performance of the RCB method degrades dramatically. Some representative alignment results are shown in Fig.5.

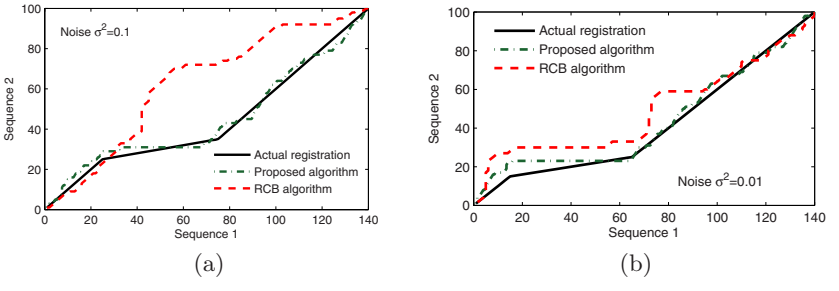


Fig. 5. Performance of the STE and RCB methods for noisy trajectories with noise variance (a) 0.1 and (b) 0.01

4.2 Real Sequences

We used video sequences provided by Rao et al. at `http://server.cs.ucf.edu/vision/` and also acquired our own video sequences of activities similar to their data. Feature trajectories were available for the UCF video files. For our test video sequences, we provided an input template image of a coffee cup that was tracked in the video sequences to generate feature trajectories. Both the proposed and RCB synchronization methods were then applied to the real video data. In

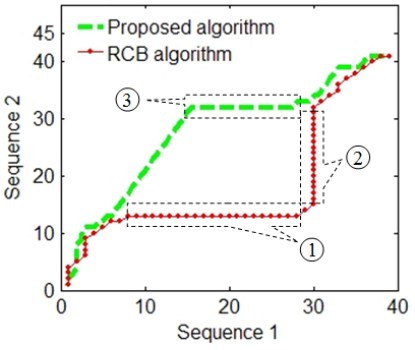


Fig. 6. Warping computed between Sequence 1 and Sequence 2 of realOUR.avi test files. Point(1)-(3) are singularities marked on the warp.

the real data tests, ground truth information is not available. However, a tentative ground truth alignment is computed by visual judgement. The test videos and results are available online at [21] as files **realUCF.avi, *realOUR.avi.*

Figure 7 shows the synchronization computed using the RCB algorithm – frames (a)-(d) of Sequence 1 are matched to frames (e)-(h) of Sequence 2. It can be seen from the position of the hand holding the cup in Fig.7 that the computed synchronization is clearly off by a few frames. Fig. 8 shows the synchronization computed by the STE algorithm, and the alignment is a close temporal match. The warping path computed between the two sequences by both the methods is shown in Fig.6. Points (1)-(3) marked on Fig.6 highlight the regions in the warp where multiple frames in Sequence 1 were warped to a single frame in Sequence 2 and vice versa —*i.e.*, singularities. It can be seen from these highlighted regions that the proposed method reduces the number and length of singularities.

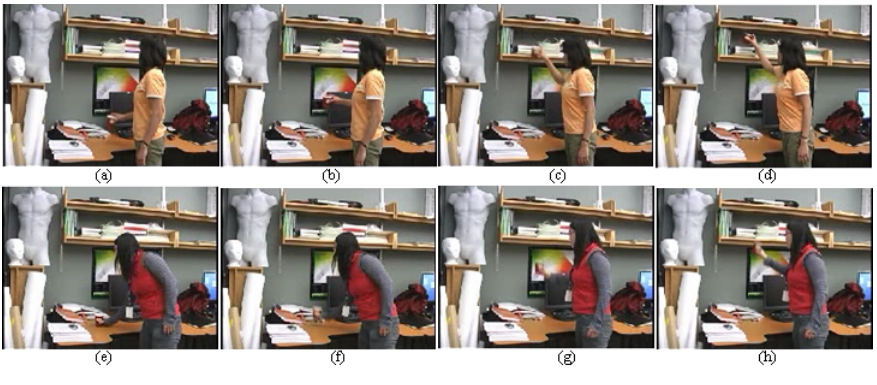


Fig. 7. Results of synchronization using a rank-constraint based RCB method. (a)-(d) Frames from Sequence 1, (e)-(h) Corresponding synchronized frames from Sequence 2.

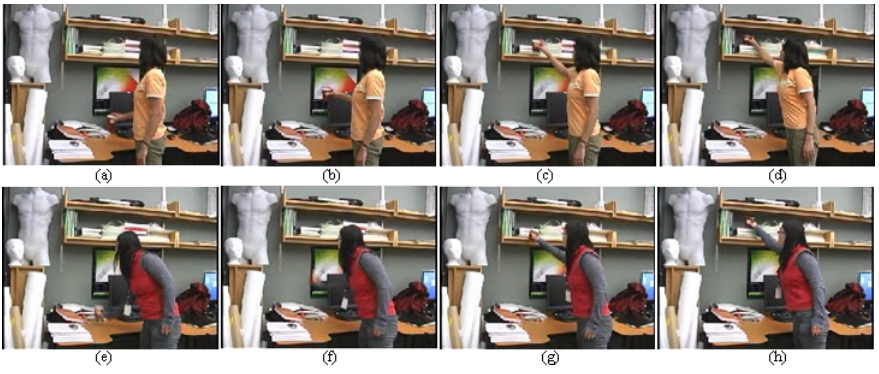


Fig. 8. Results of synchronization using proposed method. (a)-(d) Frames from Sequence 1, (e)-(h) Corresponding synchronized frames from Sequence 2.

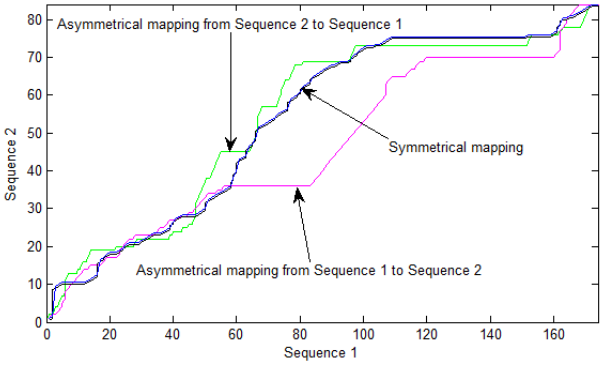


Fig. 9. Symmetric vs. Asymmetric Synchronization of realUCF video files

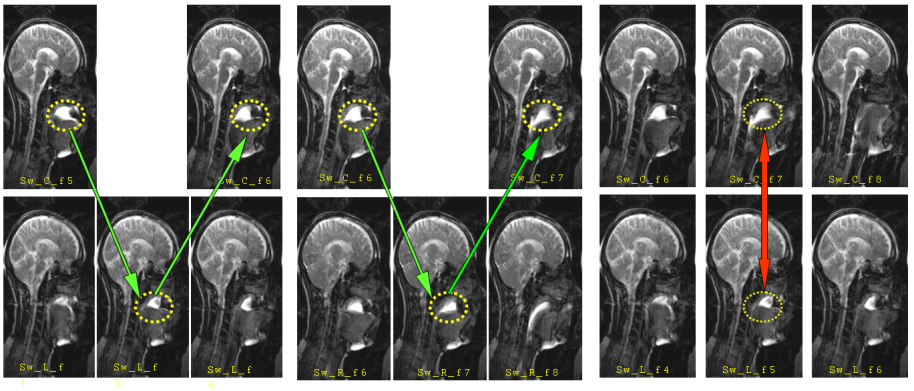
4.3 Symmetric vs. Asymmetric Synchronization

In the previous section we compared our STE based synchronization against the asymmetric synchronization proposed in [6]. The advantages of our approach in terms of sub-frame synchronization capability and reduction in singularities has been highlighted in the previously mentioned experiments. However, our approach and the RCB approach differ not only in the symmetry aspect, but also in the dynamic aspect of the cost function (4). In order to highlight the advantage of the dynamic optimization of the STE, we validate it against unidirectional asymmetric synchronization using a fixed warping cost function. The mappings computed using both the symmetric and asymmetric approach for realUCF.avi test video are shown in Fig.9. The synchronized output video files using both these approaches are available online at [21] as SymVsAsym_realUCF.avi. From the video files, the higher accuracy of a symmetric optimization over an asymmetric measure can be clearly seen.

5 Application – 4D MRI Registration

One of the motivations behind this work is to build a 4D (volume+time) representation of functional events in the body using 2D planar acquisitions, specifically swallowing disorders. In our experiments, a subject lies prone inside an MRI scanner and is fed small measured quantities of water (bolus) via a system of tubing (water is displayed as white in the MRI images because of the high Hydrogen content in it). For each swallow, a time series of 2D images is acquired on a fixed plane. The acquisition plane is then changed and another series of 2D images is acquired. We acquire three such video sequences corresponding to left, right and center MRI slice planes. The MRI video sequences are subjected to dynamic temporal offsets in the motion of the bolus. Also, since the acquisitions are at very low frame rates, limited by the technology to 4-7 fps, it is crucial to align the sequences to sub-frame accuracy. An important note, with regards to the repeatability of the swallowing motion, is that there are three stages in swallowing; the second and third stages are involuntary and anatomical movements are replicable, the first stage, when the tongue begins to push the bolus back into the pharynx is the only voluntary stage where variation in deformation is likely to take place. However, variability in the first stage can be satisfactorily controlled by limiting the volume of the bolus delivered to the subject (we limit it to 10cc) and by restricting the position of the subject (in our case - supine position with no head rotation). Our approach shows promising results in aligning the MRI sequences to generate a 4D representation. Implementation details of this application are discussed next.

The trailing and leading edges of the bolus are extracted from the MRI sequences using standard background separation techniques. The center of the trailing bolus is extracted using horizontal and vertical profiles, and is used to generate feature trajectories in the three sequences. After suitable event models have been computed for the trajectories, both the RCB and the STE algorithms are applied to the video sequences to compute synchronization. The results of synchronization with the STE and RCB algorithms are shown in Fig.10(a)-(b) and Fig.10(c) respectively. Fig.10(b) shows a few frames from the synchronization computed between the center and right MRI slices that demonstrate sub-frame alignment. Frame 7 of the right MRI sequence is mapped to frame 6.5 of the center MRI sequence. Visually it can be seen that this sub-frame alignment is quite accurate and the results are much better than those produced with the RCB algorithm shown in Fig.10(c). Video sequences comparing the results of both the algorithms are available online at [21] as files *MRI_sync*.avi*. Once the synchronization has been computed, 4D visualization of the MRI data is carried out with a 4D model which can be viewed online at [21]. We have also implemented an approach for registering multiple MRI video contours with 3D facial structure for an integrated 4D structure-multiple video visualization. A video of the implementation is also available online. Details on this part of the work will not be discussed here because of space limitations.



(a) (b) (c)

Fig. 10. (a)-(b) Synchronization computed between Center-Left and Center-Right MRI sequences respectively by proposed algorithm, (c) Synchronization computed between Center-Left MRI sequences by RCB algorithm

6 Summary and Conclusions

We proposed and successfully tested a novel method to synchronize video sequences that are related by varying temporal offsets. Our formulation of synchronization as the minimization of a symmetric transfer error (STE) resulted in synchronization that was not biased by the choice of the reference sequence. The regularized nature of the STE significantly reduced the occurrence of singularities and resulted in sub-frame synchronization. Comparative analysis with a rank-constraint based method demonstrated a marked improvement in video synchronization with our method. An application of the proposed method in 4D MRI visualization was also presented. In the future, we would like to incorporate multiple feature trajectories and extend this work to more general scenes that are not restricted by a homography relationship.

References

1. Caspi, Y., Irani, M.: Spatio-temporal alignment of sequences. PAMI 24, 1409–1424 (2002)
2. Caspi, Y., Simakov, D., Irani, M.: Feature-based sequence-to-sequence matching. IJCV 68, 53–64 (2006)
3. Singh, M., Basu, A., Mandal, M.: Event dynamics based temporal registration. IEEE Transactions on Multimedia 9, 1004–1015 (2007)
4. Lee, L., Romano, R., Stein, G.: Monitoring activities from multiple video streams: establishing a common coordinate frame. IEEE Transactions on Pattern Analysis and Machine Intelligence 22, 758–767 (2000)
5. Hess, R., Fern, A.: Improved video registration using non-distinctive local image features. In: CVPR, pp. 1–8 (2007)

6. Rao, C., Gritai, A., Shah, M., Mahmood, T.F.S.: View-invariant alignment and matching of video sequences. In: Proc. ICCV, pp. 939–945 (2003)
7. Rao, C., Shah, M., Syeda-Mahmood, T.: Invariance in motion analysis of videos. In: Proceedings of the ACM Int. Conf. on Multimedia, pp. 518–527 (2003)
8. Keogh, E., Pazzani, M.: Derivative dynamic time warping. In: SIAM International Conference on Data Mining (2001)
9. Dai, C., Zheng, Y., Li, X.: Subframe video synchronization via 3d phase correlation. In: ICIP, pp. 501–504 (2006)
10. Tuytelaars, T., VanGool, L.J.: Synchronizing video sequences. In: CVPR, pp. I: 762–768 (2004)
11. Tresadern, P., Reid, I.: Synchronizing image sequences of non-rigid objects. In: Proc. BMVC, vol. 2, pp. 629–638 (2003)
12. Carceroni, R.L., Padua, F.L.C., Santos, G.A.M.R., Kutulakos, K.N.: Linear sequence-to-sequence alignment. CVPR, I: 746–753 (2004)
13. Lei, C., Yang, Y.H.: Tri-focal tensor-based multiple video synchronization with subframe optimization. IEEE Trans. on IP 15, 2473–2480 (2006)
14. Wolf, L., Zomet, A.: Wide baseline matching between unsynchronized video sequences. IJCV 68, 43–52 (2006)
15. Pooley, D.W., Brooks, M.J., van den Hengel, A.J., Chojnacki, W.: A voting scheme for estimating the synchrony of moving-camera videos. In: ICIP, vol. 1, pp. I–413–16 (September 14-17, 2003)
16. Perperidis, D., Mohiaddin, R., Rueckert, D.: Spatio-temporal free-form registration of cardiac mr image sequences. Medical Image Analysis 9, 441–456 (2005)
17. Giese, M.A., Poggio, T.: Morphable models for the analysis and synthesis of complex motion patterns. Int. J. Comput. Vision 38, 59–73 (2000)
18. Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Gool, L.V.: A comparison of affine region detectors. Int. J. Comput. Vision 65, 43–72 (2005)
19. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide baseline stereo from maximally stable extremal regions. In: Proceedings of British Machine Vision Conference, September 2002, vol. I, pp. 384–393 (2002)
20. Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision, 2nd edn. Cambridge University Press, Cambridge (2004)
21. Authors, `www.ece.ualberta.ca/~meghna/ECCV08.html`