High-speed Motion Analysis with Multi-Exposure Images

Christian Linz, Timo Stich, Marcus Magnor

Computer Graphics Lab, TU Braunschweig
Email: {linz,stich,magnor}@cg.tu-bs.de

Abstract

We propose a method to estimate dense motion vector fields from multi-exposure images. Our approach relies on finding a sparse set of correspondences between features in a single-exposure image and each exposure in a multi-exposure image using a global optimization technique. We iteratively establish such matches, compute a set of locally restricted transformations for the matches, and construct a dense motion vector field in a multi-resolution framework. The estimation of the number of necessary transformations and the regions of influence is guided by superpixel segmentation of the image. We present results for multi-exposure photos of different dynamic scenes.

Figure 1: Multi-exposure image of an actor waving an arm.

1 Introduction

Photographic capture of high-speed motion fascinates artists and researchers alike. First investigations date back to 1878 when Eadweard Muybridge conducted his famous experiments to create serial images of a galloping horse. In the 1930's, Harold E. Edgerton perfected the use of stroboscope photography to create multi-exposure images of high-speed natural phenomena. Multi-exposure images collapse several consecutive images into one common image plane. On that account, such images implicitly convey a lot of information about the ongoing motion. Most intriguingly, our human visual system is able to reconstruct the underlying motion pattern from such multiply exposed images.

Stroboscopic photography offers a way to visualize and analyze a wide range of high-speed phenomena. For example, it would be possible to record time-varying phenomena such as an explosion and to measure the trajectory of the particles over time. Potential applications of this technique include athletics, to assist athletes in training, e.g., for a tennis player bringing his serve to perfection. Up to now, such application fields must rely on specialized and expensive high-speed cameras. With this work, we intend to demonstrate that multi-exposure images can offer a low-cost alternative to specialized hardware. So far, multi-exposure imaging has been used predominantly in situations where it was possible to augment the captured scene with markers that can be robustly and individually identified [18]. In our work, we propose a method for multi-exposure motion analysis which is completely passive. Our method requires no scene intrusion such as markers or other preparation. We only require the motion to be directed, continuous and smooth, i.e., the temporal ordering of the exposures has to map to a spatial ordering in the image plane. The input to our algorithm is the multi-exposure image to be analyzed, and a single-exposure image taken shortly before (or after) the multi-exposure image. This single-exposure is needed to initialize our iterative algorithm and to determine the direction of time. Our method is based on deformable shape matching, followed by a step to estimate a set of locally restricted aligning transformations. This set of transformations maps the initial exposure onto each exposure present in

the multi-exposure image.

As main contribution of this paper, we propose an algorithm for deducing dense motion vector fields from multi-exposure images. We incorporate Euclidean distance and shape context distance into a distance measure and use it for shape matching in cluttered environments. For increased robustness, we employ our algorithm in a multi-resolution framework. In addition, we propose a method to automatically find a set of piecewise local transformations that minimize the matching error for a given set of point correspondences.

The rest of the paper is organized as follows. In Sect. 2, we review the current state of the art. Sect. 3 discusses the representation of the scene in both single- and multi-exposure images, used by our algorithm described in Sect. 4. Finally, we present and discuss the results of the proposed algorithm for several scenes, Sect. 5.

2 Related Work

In general, automatically deducing motion vector fields from multi-exposed images is an ill-posed problem. Only for specific cases multi-exposure imaging has been used, e.g., to analyze the trajectory of a flying baseball and the hand posture of a pitcher [18]. In order to be able to analyze the motion, both the ball and prominent positions on the hand of the pitcher were augmented with markers that could be easily identified and tracked.

Shape recognition based on spatial configurations of a small number of key points, on the other hand, is a well-researched field. Belongie et al. [1] introduced the shape context descriptor which characterizes a particular point on the shape. In essence, it is a log-polar histogram of the relative coordinates of all other points. Similar points in two shapes will have a similar relative position in each shape and will ideally have a similar shape context. Shape context matching has been applied to a wide variety of object recognition tasks [1, 13] where the background clutter is limited. Thayananthan et al. propose an algorithm for shape matching based on shape contexts that is applicable also in cluttered environments [17]. They propose to integrate figural continuity into the matching framework by imposing an ordering on the object contour and penalizing matches that violate that ordering.

A standard method for point registration based on Euclidean distance is the Iterated Closest Point (ICP) algorithm [4]. Correspondences are found based on inter-point distance. The transformation is estimated by minimizing the geometric error between point pairs. This algorithm is fast and converges to a local optimum. However, it requires a good initial alignment of model and target shape. Another improved approach is the non-rigid point matching of Chui et al. and Gold et al. [6, 10] based on thin-plate spline interpolation [5]. In this work, the authors jointly estimate correspondences and a non-rigid transformation aligning the point sets. This approach has proven convergence properties and can be extended to multiple transformations, given their spatial support is known a-priori.

Concerning the estimation of dense deformation fields from a set of point correspondences, the most prominent work is the one of Bookstein [5]. In this work, the field is computed using thin-plate splines and radial basis functions for each point in the correspondence map. Recently, Schaefer et al. [16] proposed an algorithm to derive dense motion fields based on point correspondences and moving least squares interpolation. Schaefer et al. incorporate affine, similarity and rigid transformations into a common framework. Both methods yield dense and globally smooth deformation fields.

None of the mentioned shape-matching methods can directly be applied to multi-exposure images to find correspondences among overlapping scene parts. Euclidean distance-based methods such as ICP work well if the shapes are already coarsely aligned but produce wrong results if this is not the case. Shape contexts are the method of choice for matching if the amount of clutter and distortion is expected to be small. In this paper, we propose to combine the advantages of distance-based matching and shape context-based matching in a common framework, giving more weight to the Euclidean distance in already aligned areas and relying on shape contexts in areas where an alignment is not given. In addition, we propose a method for constructing discontinuity-preserving dense deformation fields from a set of sparse correspondences.

3 Shape representation

As input, we assume a multi-exposure image I_M containing k instances of an non-rigidly deforming object in motion and a single-exposure image I_S of

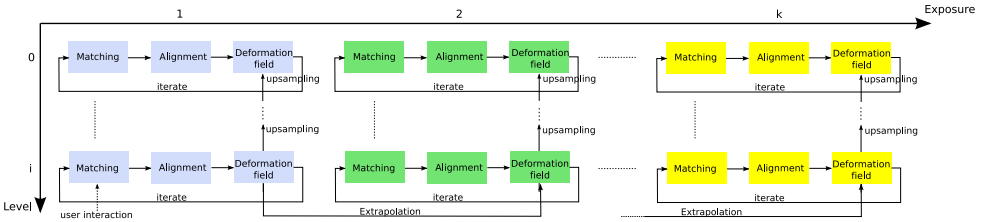
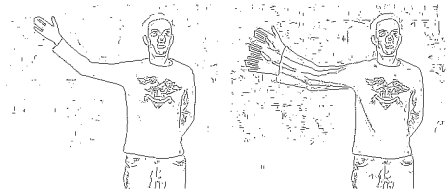
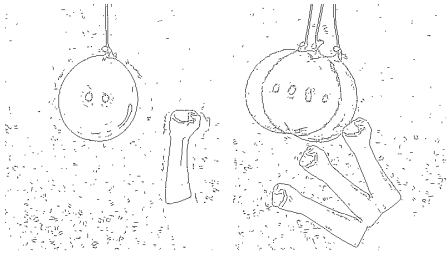


Figure 3: Schematic overview of the multi-exposure analysis algorithm. It starts at a low resolution level and iterates until convergence for the first exposure on the highest resolution level is achieved. We then extrapolate the resulting deformation to initialize our algorithm for the second exposure and start over again. This process is repeated until convergence for all k exposures is reached.



(a) Model shape, waving arm sequence

(b) Target shape, waving arm sequence



(c) Model shape, punch sequence

(d) Target shape, punch sequence

Figure 2: Shape representation for the examples provided in this paper. Note the heavy clutter resulting from overlapping exposures.

the same object before the motion sequence. Our task is to identify in I_M each of the k instances of the object depicted in I_S and compute aligning transformations that transform the single exposure image onto each of these instances. We assume linear superposition of the k instances in I_M, i.e., no saturation of pixel intensities due to multi-exposure.

Multi-exposure images are difficult to handle from an algorithmic point of view. Matching strategies based on color correlation or other intensity-based measures will only work for multi-exposure images without any overlap between consecutive exposures. In case of overlapping exposures, color information alone does not provide enough constraints to solve for the desired transformations. Since we wish to explicitly allow for overlapping exposures, we consider only the shape of the object, described by its contour. Contrary to object color/intensity, the contour is not corrupted during the multi-exposure image formation process.

We describe the object depicted in image I_S as a set of n discrete contour points $\mathcal{P} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, and the collection of exposures in I_M as the set of m points $\hat{\mathcal{P}} = \{\hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_m\}$, $\mathbf{x}_i, \hat{\mathbf{x}}_j \in \mathbb{R}^2$. Note that the points $\mathbf{x}_i$ are not required to be key points such as those found by a corner detector, nor need they be extrema of a scale space operator such as SIFT features [12]. It is sufficient to sample the points from the output of an edge detector, in our case the result of the Compass operator [15]. Furthermore, each contour point is assigned a local shape context h_i [1], a coarse histogram of relative coordinates of the neighboring points. In the following, we call the tuple $\mathbf{e}_i = (\mathbf{x}_i, h_i)$ of a contour point location $\mathbf{x}_i$ and an associated local shape context h_i an *edglet*. We further refer to the set $\mathcal{E} = \{\mathbf{e}_i | i = 1, \ldots, n\}$ as the *model shape* and the set $\hat{\mathcal{E}} = \{\hat{\mathbf{e}}_j | j = 1, \ldots, m\}$ as the *target shape*, see Fig. 2 for the examples used in this paper. Our model is similar in spirit to Active Shape Models proposed by Cootes et al. [7]. However, our approach does not rely on a statistical shape representation to restrict the deformation of the model.

As already pointed out by Thayananthan et al. [17], shape contexts become unreliable in cluttered environments. Multi-exposure images suffer from a lot of clutter, introduced by the different exposures of the object. Furthermore, we expect the

model shape to deform over a wide range, leading to completely different shape contexts for corresponding model points. Global shape contexts that take the entire image plane into account are therefore not robust for matching multi-exposure images. Instead, we use a very localized shape context, computed from a dozen neighboring points around each point. This minimizes the corruption of the shape contexts in heavily overlapping regions, while it still offers a reasonable expressiveness for non-overlapping parts of the shape.

4 Algorithm

Our multi-exposure analysis algorithm consists of three main parts: correspondence estimation, computation of a set of aligning transformations, and construction of a dense deformation field. These three steps are repeatedly iterated until convergence for a certain exposure is reached. Inspired by coarse-to-fine optical flow proposed by Bergen et al. [2], we employ our algorithm in a multi-resolution framework, i.e., we apply it to an image pyramid starting at low resolution. This way, we are able to quickly find a coarse alignment of the model shape and a given exposure of the target shape, which is a vital condition for distance-based matching at higher resolutions. Results are transferred to the next higher level in the hierarchy by up-sampling the deformation field. The algorithm is repeated until all exposures are matched. A schematic overview is given in Fig. 3. The subroutines are explained in the following subsections.

4.1 Matching shapes

Given a model and a target shape, we first need to find a mapping Φ from the set of model shape edglets $\mathcal{E}$ to the set of target shape edglets $\hat{\mathcal{E}}$, i.e., $\Phi : \mathcal{E} \rightarrow \hat{\mathcal{E}}$. We require the mapping Φ to be one-to-one. This is the well-known assignment problem which can be encoded in a graph structure [8].

In our setup, the graph consists of $2n + m$ nodes, Fig. 4; one node for each edglet in $\mathcal{E}$, one node for each edglet in $\hat{\mathcal{E}}$, and one additional occluder node for each edglet of the model shape. The occluder nodes allow the matching to be one-to-one without the need to match an actually existing edglet, making it possible to match edglet sets of unequal sizes. For each node of the model shape, we add a graph

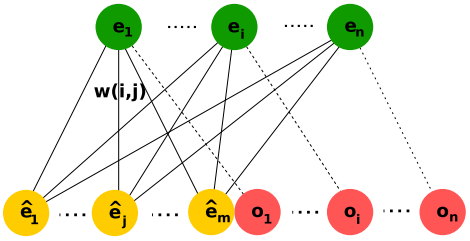


Figure 4: The assignment graph. Each node of the model shape (green) is connected to each node of the target shape (yellow). Furthermore, each node of the model shape is connected to its occluder node (red).

edge to each node of the target shape. Furthermore, each model shape node is connected to its occluder node, Fig. 4. Finally, each edge between a model and a target node is assigned a weight, modeled by the following formula

$$w(\mathbf{e}_i, \hat{\mathbf{e}}_j) = -(1 + \delta(d_{ij}, \chi^2_{ij})) \cdot D(d_{ij}) \quad (1)$$

where

$$\begin{aligned} d_{ij} &= \|\mathbf{A}_l \mathbf{x}_i - \hat{\mathbf{x}}_j\|^2, \\ \chi^2_{ij} &= \frac{1}{2} \sum_k \frac{[h_i(k) - \hat{h}_j(k)]^2}{h_i(k) + \hat{h}_j(k)}, \\ D(x) &= \frac{a}{1 + e^{-bx}}, \\ \delta(x, \chi^2) &= \begin{cases} 0, x \leq c \\ \chi^2, \text{otherwise} \end{cases}. \end{aligned}$$

a and b are chosen such that the maximal cost for the Euclidean distance is limited by a. c is found empirically and is set to 20. $\mathbf{A}_l$ is the locally restricted aligning transformation for translet l, cf. Sect. 4.2, and is initialized to the identity matrix. The edges to the occluder nodes are treated separately and are assigned a fixed weight. This weight models the maximal feasible matching distance and can be used to restrict the matching. By combining shape context distance and Euclidean distance into the weight function (1), we combine the advantages of both matching approaches. When the distance between edglets is smaller than the threshold c, the influence of the shape context is discarded, i.e. the matching is driven by Euclidean distance. Vice versa, if the distance is larger, shape context similarity becomes important and helps to distinguish

wrong matches from correct ones. Small distances between matching edglets usually occur in heavily overlapping regions, regions where shape context-based matching is problematic; larger displacements potentially lead to less overlap and hence to an increase in the reliability of localized shape contexts. Furthermore, small displacements also mean that parts of the shape are already coarsely aligned, further justifying a matching based on Euclidean distance.

The solution to the assignment problem we are searching for maximizes the benefit over all assignments $\mathcal{S}$,

$$\max_{s \in \mathcal{S}} \sum_{i=1}^{n} w(\mathbf{e}_i, \Phi_s(\mathbf{e}_i)).$$

An optimal assignment is then found by applying the auction algorithm [3] to the graph constructed above. The auction algorithm is used to solve the assignment problem since its time complexity [3] ($\mathcal{O}(nA\log(nC))$) is better than that of the Hungarian method [14] ($\mathcal{O}(n^3)$) which is usually used to solve this type of problem. Pairwise matching methods such as spectral matching [11] are not applicable in this case given the sheer number of edglets to be matched. While the assignment found by the algorithm is optimal in terms of the weight function (1), the resulting matching still contains outliers. However, we do not aim at producing optimal matches (considered from the point of view of a human observer) within one iteration of the matching algorithm. Instead, we apply several iterations of the algorithm, interwoven with the estimation of a set of aligning transformations described in the following section, Fig. 3. Furthermore, the user initially has the possibility to influence the matching by defining preferred edglet matches.

4.2 Aligning shapes

Given a mapping Φ, we now want to estimate a set of aligning transformations that map each feature to its found correspondence. Methods proposed in the literature such a thin-plate spline interpolation [5] or moving-least squares based interpolation [16] try to find a globally smooth transformation for the given set of correspondences. However, a globally smooth solution is too restrictive, e.g., those methods offer no way to account for discontinuities and can be prohibitively expensive to compute given more than a thousand feature pairs in our setup [16].

Instead, our goal is to construct piecewise local transformations that only map a certain subset of edglet correspondences onto each other and that allow to preserve discontinuities in the field. To this end, we first have to decide how many transformations are needed to faithfully align the shapes. In order to start with a good guess for the number of transformations, we compute a superpixel segmentation [9] of the single-exposure image I_S. This yields a conservative over-segmentation of the image where each segment in turn is a consistent unit, i.e, all pixels in a superpixel are most similar in color and texture and are thus likely to move under the same transformation. The size of a superpixel is typically in the order of twenty pixels.

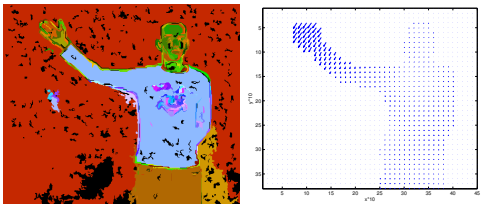
Each superpixel is then assigned the set of edglet matches that lie within its image region. We call this combination of a superpixel and a set of edglet matches a *translet* $\mathbf{t}_l$. Next, we estimate a transformation matrix $\mathbf{A}_l$ for each translet $\mathbf{t}_l$ if it contains a minimal number of edglet matches. In order to robustly estimate the transformation in the presence of potential mismatches, we apply the RANSAC algorithm to find the best transformation and to further classify the set of edglet matches into inlier and outlier matches based on the matching error

$$\epsilon_{i,j} = \|\mathbf{A}_l \mathbf{x}_i - \hat{\mathbf{x}}_j\|.$$

There is no restriction on the kind of transformation, but in our experiments affine transformations turned out to give the best results. Based on the estimated transformation, we compute an error measure for the current segmentation,

$$E_{Seg} = \sum_{l} \sum_{(\mathbf{e}_i, \hat{\mathbf{e}}_j) \in \mathbf{t}_l} \epsilon_{i,j}. \qquad (2)$$

The next step is to optimize the superpixel segmentation to better represent the distribution of local affine transformations and to minimize the accumulated error E_{Seg} in (2). Since the initial superpixel segmentation usually results in a strong over-segmentation, we are able to combine neighboring translets with similar transformations. In theory, it may sometimes be necessary to split a translet; in our experiments, however, this case rarely occured and its influence can be safely neglected. The optimization is carried out in a greedy manner: in each iteration, we find the neighboring translets that yield



(a) Translet map (b) Corresponding dense field

Figure 5: The translet map (a) encodes the region of influence of each transformation. Black areas do not belong to any translet since they contain no edglets. Based on those transformations, we construct a dense deformation field, shown in (b).

the highest improvement for (2), collapse the two and estimate a new transformation for the resulting translet. This greedy optimization is repeated until either the accumulated error reaches a minimum, or a given minimal number of translets is obtained.

At this stage, we have a set of translets, each defining a locally optimal affine transformation for the given matches. It further defines the regions of influence of each transformation. Fig. 5a shows the final translet map for the first exposure of the multi-exposure image shown in Fig. 1. Regions of the same color encode the region of influence of a translet. Black areas do not belong to any translet since they contain no edglets.

4.3 Deriving a dense deformation field

So far, the correspondences only define a sparse deformation field. Since we want to transform the entire shape, we have to construct a dense deformation field from the set of sparse correspondences. The construction should smooth small differences between neighboring translets with similar transformations, while preserving large differences. Furthermore, the resulting deformation should interpolate the estimated correspondences. To this end, we initially consider each translet in the translet map, Fig. 5a, and compute per-pixel displacement vectors using the estimated transformation. In a second step, we smooth the resulting dense field by solving an anisotropic nonlinear diffusion PDE [19] on the entire image plane. In order to enforce consistency with the estimated matches, we treat the corresponding displacement vectors as boundary conditions to the diffusion process. By applying nonlinear diffusion, we smooth out small differences while preserving discontinuities in the field, Fig. 5b.

4.4 Avoiding multiple matches in the multi-exposure shape

After the algorithm has converged for a certain exposure, we need to decide what to do with the matched edglets in the target shape. Some of them can be removed since they only belong to the current exposure, others may belong to more than one exposure, e.g., edglets belonging to a region of the scene that doesn't move. These edglets have to be kept in order to provide good alignments for the following exposures. Therefore, instead of removing edglets from the target shape, we penalize matches to already matched edglets in further iterations of our algorithm. We set the penalty to be a multiple n of the matching distance and extend the weight function given in (1) to

$$\tilde{w}(\mathbf{e}_i, \hat{\mathbf{e}}_j) = w(\mathbf{e}_i, \hat{\mathbf{e}}_j) - n \cdot \|\mathbf{x}_p - \hat{\mathbf{x}}_j\|^2, \quad (3)$$

where $\mathbf{x}_p$ is the source of the previous match for target edglet $\hat{\mathbf{x}}_j$. We extrapolate the final deformation field for the current exposure to get a good starting point for the next exposure. This also results in a speed-up since fewer iterations are necessary to converge to a stable matching.

5 Results

We present results for fast motion sequences carried out by human actors. All images are taken using a Canon EOS 5D digital camera, equipped with a 28mm prime lens. The environment has to be completely dark, the only source of illumination being a high-output stroboscope.The exposure time of the cameras is set such that several flashes are recorded and the entire motion sequence is mapped in one image. The frequency f of the stroboscope is used to control the number of exposures captured in the multi-exposure image. For the multi-exposure image shown in Fig. 6, the frequency of the stroboscope was set to $f = 45Hz$. The single-exposure image should be taken around $\frac{1}{f}$ seconds before the multi-exposure to ease extrapolation.

We present results for the multi-exposure images shown in the first row of Fig. 6. The images show the recovered deformation fields for each exposure, and their application to the single-exposure image

for two different time instants. The last row of Fig. 6 shows a synthesized motion-blurred image computed from the recovered intermediate time instants for both examples. As can be seen in Fig. 6, our method is capable of reconstructing dense motion vector fields from multi-exposure images displaying a smooth and continuous motion. The motion vector fields yield plausible results when applied to the single-exposure image. Fine details as the fingers in the waving arm sequence and the highlights on the balloon of the punch sequence are well preserved.

5.1 Limitations

Concerning the accuracy of our method, we are aware that it is not comparable to optical flow methods. Our algorithm yields correct per-pixel displacement vectors only for those pixels identified as edglets. Fine details which are not covered by the shape representation may be only coarsely aligned, for example the fingers in Fig. 6 (v).

Our method works for directed, continuous and smooth types of high-speed motion, i.e., types of motion where the temporal ordering induces a spatial ordering in the image plane. The single-exposure image used to initialize the proposed method needs to obey that ordering constraint, too. Furthermore, it also has to fit into the sampling pattern of the multi-exposure image. The latter requirement is due to the fact that we extrapolate the deformation field to get a good initialization for the next exposure.

Since our method relies on linear photometric superposition of exposures, it is restricted by the dynamic range and the sensitivity of the camera's sensor. With higher repetition rate, the duration of each flash, and hence the amount of light, decreases. With too many exposures, sensor elements start to saturate. Both effects result in a decrease in contrast in the multi-exposure image and lead to missing edges in the shape representation until the shapes can no longer be aligned. Experiments have shown that for multi-exposure images exhibiting a strong overlap, up to eight exposures can be robustly detected.

5.2 Ongoing work

Using the presented algorithm, we are able to warp the single exposure image to each exposure in the multiply exposed image. As pointed out already, the accuracy of our algorithm is not good enough in regions with sparse edge information. Given the warped images, we can now proceed to further refine each single deformation field by an analysis-by-synthesis approach. To this end, we plan to synthesize a multi-exposure image from the warped single exposures and compare it the recorded multi-exposure image. Minimizing the residual will hopefully lead to more accurate deformation fields, especially in regions without edge information.

6 Conclusion

We presented an approach to estimate dense motion vector fields from a multi-exposure image of high speed motion. We demonstrated the applicability of our method to real-world non-rigid scenes. We discussed the capabilities of our method and pointed out its current limitations.

We believe that our work is a step towards image-based analysis of fast motion events without the need for special hardware. Future work includes the development of a suitable hardware setup to simplify recording. Furthermore, we will concentrate on developing a shape representation that lifts the ordering constraint for the multi-exposure images. Finally, we intend to extend our approach to multiple views in order to estimate dense 3D motion vector fields.

References

[1] Serge Belongie, Jitendra Malik, and Jan Puzicha. Matching shapes. In *Proc. ICCV*, volume 1, pages 454–463. IEEE Computer Society, July 2001.

[2] James R. Bergen, P. Anandan, Keith J. Hanna, and Rajesh Hingorani. Hierarchical model-based motion estimation. In *Proc. ECCV*, pages 237–252, London, UK, 1992. Springer-Verlag.

[3] D. P. Bertsekas. The auction algorithm: a distributed relaxation method for the assignment problem. *Ann. Oper. Res.*, 14(1-4):105–123, 1988.

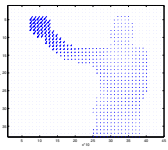
[4] Paul J. Besl and Neil D. McKay. A method for registration of 3-d shapes. *IEEE TPAMI*, 14(2):239–256, 1992.

(a) Single-exposure ($t = 0$), waving arm sequence

(b) Multi-exposure, waving arm sequence

(c) Single-exposure ($t = 0$), punch sequence

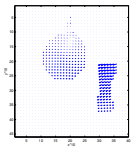
(d) Multi-exposure, punch sequence



(e) Deformation for first exposure.

(f) (a) warped to $t = 0.5$,

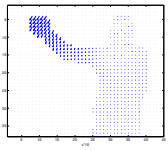
(g) and to $t = 1.0$.



(h) Deformation for first exposure.

(i) (c) warped to $t = 0.5$

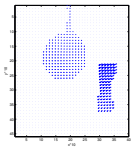
(j) and to $t = 1.0$.



(k) Deformation for second exposure.

(l) (a) warped to $t = 1.5$

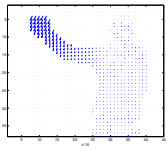
(m) and to $t = 2.0$.



(n) Deformation for second exposure.

(o) (c) warped to $t = 1.5$

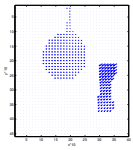
(p) and to $t = 2.0$.



(q) Deformation for third exposure.

(r) (a) warped to $t = 2.5$

(s) and to $t = 3.0$.



(t) Deformation for third exposure.

(u) (c) warped to $t = 2.5$

(v) and to $t = 3.0$.

(w) Motion-blurred image.

(x) Motion-blurred image.

Figure 6: Results for the waving arm sequence (left part) and the punch sequence (right part). The first row shows the input to our algorithm. The following three rows show the recovered deformation fields and their application to the single exposure image for the first, second and third exposure (from top to bottom). The intermediate exposures are generated by forward warping the single exposure. The last row shows a synthesized motion-blurred image, summarizing the motion for both sequences.

[5] F. L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE TPAMI*, 11(6):567–585, 1989.

[6] H. Chui and A. Rangarajan. A new algorithm for non-rigid point matching. In *Proc. CVPR*, pages 44–51. IEEE Computer Society, 2000.

[7] T. Cootes and C. J. Taylor. Active shape models - 'smart snakes. In *In British Machine Vision Conference*, pages 266–275. Springer-Verlag, 1992.

[8] Thomas H. Cormen, Charles E. Leiserson, and Ronal L. Rivest. *Introduction to Algorithms, Second edition*. The MIT Press, Cambrigde, Massachusetts, 2001.

[9] Pedro F. Felzenszwalb and Daniel P. Huttenlocher. Efficient graph-based image segmentation. *IJCV*, 59(2):167–181, 2004.

[10] S. Gold, A. Rangarajan, C.P. Lu, S. Pappu, and E. Mjolsness. New algorithms for 2d and 3d point matching: Pose estimation and correspondence. *Pattern Recognition*, 31(8):1019–1031, 1998.

[11] Marius Leordeanu and Martial Hebert. A spectral technique for correspondence problems using pairwise constraints. In *Proc. ICCV*, pages 1482–1489, Washington, DC, USA, 2005. IEEE Computer Society.

[12] David G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[13] Greg Mori and Jitendra Malik. Estimating human body configurations using shape context matching. In *Proc. ECCV (3)*, pages 666–680, 2002.

[14] J. Munkres. Algorithms for the Assignment and Transportation Problems. *Journal of the Society of Industrial and Applied Mathematics*, 5(1):32–38, 1957.

[15] Mark A. Ruzon and Carlo Tomasi. Color edge detection with the compass operator. In *Proc. CVPR*, pages 2160–2166. IEEE Computer Society, 1999.

[16] Scott Schaefer, Travis McPhail, and Joe Warren. Image deformation using moving least squares. In *Proc. SIGGRAPH'06*, pages 533–540. ACM SIGGRAPH, 2006.

[17] A. Thayananthan, B. Stenger, P.H.S. Torr, and R. Cipolla. Shape context and chamfer matching in cluttered scenes. In *Proc. CVPR*, pages 127–133. IEEE Computer Society, 2003.

[18] Ch. Theobalt, I. Albrecht, J. Haber, M. Magnor, and H.-P. Seidel. Pitching a Baseball – Tracking High-Speed Motion with Multi-Exposure Images. In *Proc. SIGGRAPH '04*, pages 540–547. ACM SIGGRAPH, 2004.

[19] Joachim Weickert. *Anisotropic Diffusion in Image Processing*. Teubner, 1998.