Multi-reference combinatorial strategy towards longer long-term dense motion estimation

Pierre-Henri Conze, Philippe Robert, Tomas Crivelli, Luce Morin

Highlights

- long-term dense motion estimation based on multi-step optical flows is addressed
- optical flows are combined through multi-step integration and statistical selection
- an analysis of available single-reference complexity reduction schemes is provided
- our new multi-reference frames processing reaches longer accurate displacement fields

Multi-reference combinatorial strategy towards longer long-term dense motion estimation

Pierre-Henri Conze[a], Philippe Robert[b], Tomás Crivelli[b], Luce Morin[c]

[a]ICube UMR 7357, Université de Strasbourg, CNRS, 300 boulevard Sébastien Brant, CS 10413, 67412 Illkirch Cedex, France
[b]Technicolor, 975 avenue des Champs Blancs CS 17616, 35576 Cesson-Sévigné, France
[c]INSA Rennes, IETR, UMR 6164, UEB, 20 avenue des Buttes de Coesmes, 35708 Rennes Cedex 7, France

Abstract

This paper addresses the estimation of accurate long-term dense motion fields from videos of complex scenes. With computer vision applications such as video editing in mind, we exploit optical flows estimated with various inter-frame distances and combine them through multi-step integration and statistical selection (MISS). In this context, managing numerous combinations of multi-step optical flows requires a complexity reduction scheme to overcome computational and memory issues. Our contribution are two-fold. First, we provide an exhaustive analysis of available single-reference complexity reduction strategies. Second, we propose a simple and efficient alternative related to multi-reference frames multi-step integration and statistical selection (MR-MISS). Our method automatically inserts intermediate reference frames once matching failures are detected to re-generate the motion estimation process and re-correlates the resulting dense trajectories. By this way, it reaches longer accurate displacement fields while efficiently reducing the complexity. Experiments on challenging sequences reveal improved results compared to state-of-the-art methods including existing MISS schemes both in terms of complexity reduction and accuracy improvement.

Keywords: long-term motion estimation, dense matching, multi-reference frames, combinatorial integration, motion trajectories, video editing

1. Introduction

Estimating accurate long-term dense correspondence fields is a fundamental task for many computer vision applications. A key tool in this context is optical flow whose early formulations come from the early 80s [1, 2]. Significant progress has been made to improve both robustness and spatial consistency of the flow by introducing respectively more sophisticated data models than the classical brightness constancy assumption and robust discontinuity-preserving smoothness constraints.

However, most of state-of-the-art optical flow estimators focus on estimating dense motion between two consecutive frames only. They seldom consider that sequences comprise series of images that are inter-related. When tackling motion estimation over a video sequence, object-based [3] or sparse [4] motion estimation is usually sufficient (visual servoing, surveillance, gestural human-machine interface, video indexing...). However, other applications explicitly require a dense and long-term description of how the video content evolves in time. Such applications include scene segmentation [5, 6], trajectory analysis [7] or video editing tasks like 2D-to-3D video conversion [8] or graphic elements insertion where each pixel of a given area needs to be tracked over many frames to be properly replaced by the corresponding pixel of the inserted element. Thus, we focus on this challenging issue: how to construct dense fields of point correspondences over extended time periods?

Establishing dense long-term correspondences requires to compute dense motion fields between distant frames and therefore to simultaneously handle small and large displacements. Optical flow is the appropriate tool for this task but classical optical flow assumptions which may fail between consecutive frames are even less valid between non-consecutive frames. When dealing with multiple frames and their associated point correspondences, another key aspect is the temporal consistency of the flow vectors which must depict temporal smoothness along trajectories. In this context, several recent studies have extended optical flow to the purpose of (semi-)dense long-term motion estimation. State-of-the-art deals with consecutive optical flow concatenation [5, 9, 10], trajectorial regularization [11, 12], particle representation [13], subspace constraints [14, 15] as well as multi-step strategies [16, 17, 18, 19].

Optical flows estimated between consecutive frames can straightforwardly be concatenated to construct motion trajectories along a video sequence through Euler or Runge-Kutta temporal integration [20]. This strategy has been exploited in many works [5, 9, 10] but may lead to large error accumulation resulting in a substantial drift over extended periods of time. Results in the literature are generally reported on fairly short sequences and reliable tracks usually do not exceed thirty frames.

To limit motion drift, optical flow estimation has been extended from two frames to multiple frames via hard [11] or soft [12] spatio-temporal constraints which penalize motion variations along trajectories. Despite these contributions, more sophisticated motion models have been investigated to deal with complex motion. In [13], Sand and Teller represent video motion using a set of particles that move across the sequence. To reach variable-length point trajectories, particles are sequentially propagated using optical flows computed between consecutive frames. Using such representation forsakes rigidity as-

sumptions and motion model considerations which may fail in complex situations but does not achieve full density.

Since trajectories of points belonging to an object are correlated even with strong deformations, subspace constraints-based methods assume that the set of all flow fields reside in a low-dimensional subspace [14]. Therefore, a low-rank space is built to constrain optical flow estimation which provides additional information to solve the ambiguity in regions that suffer from the aperture problem. In [15], Garg et al. perform dense multi-frame optical flow estimation in a variational framework using 2D trajectory subspace constraints [15]. This approach generates dense trajectories starting from a reference frame in a non-rigid context. Trajectories are estimated close to a low-dimensional trajectory subspace built through Principal Component Analysis (PCA) or Discrete Cosine Transforms (DCT). Nevertheless, this method requires strong a priori assumptions on the scene content. Moreover, only trajectories starting from a fixed reference frame are considered. The computation of motion fields starting from subsequent frames and going back to the reference frame is not under consideration.

The alternative concept of multi-step flow (MSF) [16, 17] focuses on how to construct long-term dense fields of correspondences using multi-step optical flows, i.e. optical flows computed between consecutive frames or with larger inter-frame distances. MSF sequentially merges a set of displacement fields at each intermediate frame, up to the target frame. This set is obtained via concatenation of multi-step optical flows with displacement vectors already computed for neighbouring frames. Multi-step estimations can handle temporary occlusions since they can jump occluding objects. Contrary to [15], MSF considers both trajectory estimation between a reference frame and all the images of the sequence (from-the-reference) and motion estimation to match each image to the reference frame (to-the-reference). Two set-ups can be then considered: information pushing from the reference frame or information propagation over each frame by pulling it from the reference frame.

Despite its ability to handle both scenarios, MSF has two main drawbacks. First, it performs the selection of displacement fields by relying only on classical optical flow assumptions such as the brightness constancy constraint that may fail between distant frames. Second, the candidate displacement fields are based on previous estimations. It ensures a certain temporal consistency but can also propagate estimation errors along the subsequent frames of the sequence, until a new available step gives a chance to match with a correct location again.

These limitations can be solved by considering the multi-step integration and statistical selection (MISS) introduced in [18, 19] for the estimation of from-the-reference and to-the-reference long-term dense motion correspondences between a reference frame I_{ref} and all the other frames I_n of a video sequence. Based on pre-computed multi-step optical flows, similarly to MSF [17], MISS algorithm processes each pair of frames $\{I_{ref}, I_n\}$ via both multi-step integration and statistical selection. Multi-step integration builds a large set of candidate displacement fields via the generation of multiple motion paths made of concatenated multi-step optical flows. Then, the statistical selection consists in selecting among the resulting set of candidate displacement fields the optimal one based on statistics and spatial regularization.

The statistical selection performs the displacement field selection by studying the redundancy on the large candidate set resulting from multi-step integration. For distant frames, it provides a more robust indication than classical optical flows assumptions involved in MSF [17]. Moreover, contrary to MSF [17] which sequentially relies on previously established correspondences, MISS algorithms independently process each pair of frames $\{I_{ref}, I_n\}$ to prevent error propagation. Temporal consistency is handled a-posteriori through robust temporal smoothness constraints [19].

Each time the multi-step integration stage processes a given pair $\{I_{ref}, I_n\}$, only a subset of all the possible motion paths between I_{ref} and I_n can be generated and kept in memory due to computational and memory issues. For instance, the number of possible motion paths for a distance of 30 frames and with steps 1, 2, 5 and 10 is... 5877241! Up to a few hundreds can be actually built and kept in memory with current computer capabilities. To avoid these issues, the multi-step integration stage must include a computational complexity reduction strategy to prevent a cumbersome exhaustive motion paths generation process. This complexity reduction scheme must cleverly select a subset of all possible motion paths to minimize the tracking failure probability while increasing the trajectory lifetime.

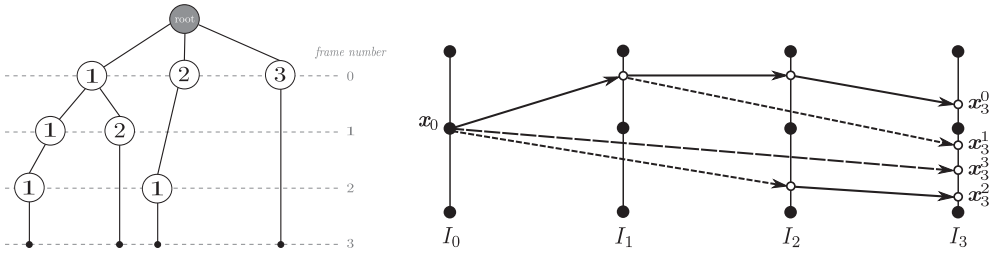
In this direction, we aim at covering and extending the spectrum of MISS introduced in [18, 19] in the context of long-term dense motion estimation. After a brief overview of the baseline method (Sect.2), two main contributions are addressed. First, given the computational and memory issues mentioned above, we identify and study the available single-reference complexity reduction schemes adapted to the multi-step integration stage of MISS (Sect.3). Second, we propose a new, simple and efficient complexity reduction strategy based on an automatic multi-reference frames processing (Sect.4). It reaches longer accurate displacement fields while efficiently reducing the complexity. Its ability to go towards longer long-term dense motion estimation is assessed through comparisons with state-of-the-art methods on challenging sequences (Sect.5).

2. Multi-step integration and statistical selection (MISS)

The baseline multi-step integration and statistical selection (MISS) method [18, 19] can be, at first glance, studied without any complexity reduction considerations. Let us overview both multi-step integration and statistical selection steps in the context of exhaustive motion path generation. For the sake of clarity, complexity reduction is addressed only starting from Sect.3.

2.1. Multi-step integration

The multi-step integration aims at producing a large set of displacement fields between a reference frame I_{ref} and a given subsequent frame I_n as to form a significative set of samples upon which a statistical processing would be meaningful and advantageous. As inputs, it takes a set of optical flow fields pre-estimated from each frame of the sequence including I_{ref}.



(a) step sequences (b) motion paths

Figure 1: Multi-step integration: (a) Generation of step sequences from I_0 to I_3 with steps 1, 2, and 3 by creating a tree structure: $\mathbf{\Gamma}_{0,3} = \{\{1,1,1\},\{1,2\},\{2,1\},\{3\}\}$; (b) Generation of motion paths following all the step sequences of $\mathbf{\Gamma}_{0,3}$ which gives for each pixel $\boldsymbol{x}_0$ of I_0 a set of candidate positions in I_3: $T_{0,3}(\boldsymbol{x}_0) = \{\boldsymbol{x}_0^3, \boldsymbol{x}_1^3, \boldsymbol{x}_2^3, \boldsymbol{x}_3^3\}$.

These optical flows are computed between consecutive frames or with larger steps [16], i.e. larger inter-frame distances. Let $S_n = \{s_1, s_2, \ldots, s_{Q_n}\} \subset \{1, \ldots, N-n\}$ be the set of Q_n possible steps at instant n. The following set of optical flow fields starting from I_n is therefore available: $\{\boldsymbol{v}_{n,n+s_1}, \boldsymbol{v}_{n,n+s_2}, \ldots, \boldsymbol{v}_{n,n+s_{Q_n}}\}$.

The starting point of multi-step integration consists in initially generating all the possible step sequences, i.e. combinations of steps, in order to join I_n from I_{ref}. Each generated step sequence defines a motion path which links each grid point $\boldsymbol{x}_{ref}$ of I_{ref} to a non-necessary grid position in I_n through multiple concatenations of un-occluded multi-step optical flow fields.

Let $\mathbf{\Gamma}_{ref,n} = \{\boldsymbol{\gamma}_0, \boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_{K-1}\}$ be the set of K possible step sequences $\boldsymbol{\gamma}_i$ between I_{ref} and I_n. A step sequence $\boldsymbol{\gamma}_i = \{s_1^i, s_2^i, \ldots, s_{K_{\boldsymbol{\gamma}_i}}^i\}$ is defined by a set of $K_{\boldsymbol{\gamma}_i}$ steps s_k^i which once cascaded join I_n from I_{ref}. The set of K possible step sequences $\mathbf{\Gamma}_{ref,n}$ is computed by building a tree structure (Fig.1a) where each node corresponds to an optical flow field assigned to a given frame for a given step value, the node value. Going from the root node to leaf nodes of this tree structure gives $\mathbf{\Gamma}_{ref,n}$, the set of K possible step sequences from I_{ref} to I_n.

Once all the possible step sequences $\boldsymbol{\gamma}_i \ \forall i \in [\![0, \ldots, K-1]\!]$ between I_{ref} and I_n are generated, the corresponding motion paths are constructed through motion vector concatenation. Starting from each pixel $\boldsymbol{x}_{ref} \in I_{ref}$ and for each step sequence $\boldsymbol{\gamma}_i$, this integration performs the accumulation of optical flow fields following the steps which form the current step sequence, i.e. $s_1^i, s_2^i, \ldots, s_{K_{\boldsymbol{\gamma}_i}}^i$ (Fig.1b). Let $f_j^i = ref + \sum_{k=1}^{j} s_k^i$ be the current frame number during the construction of motion path i from I_{ref} where j is the step index within the step sequence $\boldsymbol{\gamma}_i$. For each $\boldsymbol{\gamma}_i \in \mathbf{\Gamma}_{ref,n}$ and for each step $s_j^i \in \boldsymbol{\gamma}_i$, the integration starts from $\boldsymbol{x}_{ref}$ to iteratively compute the successive positions of motion path i along the sequence:

$$\boldsymbol{x}_{f_j^i}^i = \boldsymbol{x}_{f_{j-1}^i}^i + \boldsymbol{v}_{f_{j-1}^i, f_j^i}(\boldsymbol{x}_{f_{j-1}^i}^i) \quad (1)$$

Once all the steps $s_j^i \in \boldsymbol{\gamma}_i$ have been run through, one gets $\boldsymbol{x}_n^i$, the corresponding position in I_n of $\boldsymbol{x}_{ref} \in I_{ref}$ obtained with step sequence $\boldsymbol{\gamma}_i$. A large set of candidate positions in I_n is finally reached by considering all the step sequences of $\mathbf{\Gamma}_{ref,n}$ (Fig.1b) and this for each pixel $\boldsymbol{x}_{ref} \in I_{ref}$. Thus, to each pixel $\boldsymbol{x}_{ref}$ of I_{ref} is associated a large set of candidate positions in I_n defined as $T_{ref,n}(\boldsymbol{x}_{ref}) = \{\boldsymbol{x}_n^i\} \ \forall i \in [\![0, \ldots, K_{\boldsymbol{x}_{ref}} - 1]\!]$ where $K_{\boldsymbol{x}_{ref}}$ is the cardinal of $T_{ref,n}(\boldsymbol{x}_{ref})$.

2.2. Statistical selection

The statistical selection aims at selecting the optimal candidate position $\boldsymbol{x}_n^*$ in $T_{ref,n}(\boldsymbol{x}_{ref}) = \{\boldsymbol{x}_n^i\}_{i \in [\![0,\ldots,K_{\boldsymbol{x}_{ref}}-1]\!]}$, the set of candidate positions in I_n obtained for each pixel $\boldsymbol{x}_{ref}$ of I_{ref}. The selection of the optimal candidate position is performed by combining a statistical processing applied for each pixel $\boldsymbol{x}_{ref}$ independently as well as a global optimization method introducing spatial regularization into the candidate selection process.

For each $\boldsymbol{x}_{ref} \in I_{ref}$, the statistical processing is applied to $T_{ref,n}(\boldsymbol{x}_{ref})$ to select the N_{opt} best candidates of the distribution based on spatial density and intrinsic candidate quality. Then, the fusion moves algorithm citelempitsky2010fusion fuses the resulting N_{opt} dense candidate displacement fields pair by pair up to obtain the optimal displacement field between the distant frames I_{ref} and I_n. These fusions are performed via global optimization [18] and involve both matching cost and inconsistency quality features described in Appendix B.

The key aspect of the statistical selection relies on the selection of the N_{opt} optimal candidate positions through statistical processing. To select these N_{opt} candidates, it exploits the statistical information on the point distribution as well as information relative to intrinsic candidate quality. Based on the Maximum Likelihood Estimator (MLE) estimator for which the mean operator has been replaced by the median operator to be more robust to outliers, the choice of the N_{opt} optimal candidate positions in $T_{ref,n}(\boldsymbol{x}_{ref})$ is recursively performed following:

$$\boldsymbol{x}_n^* = arg \min_{\boldsymbol{x}_n^i} med_{j \neq i} \left\| \boldsymbol{x}_n^j - \boldsymbol{x}_n^i \right\|_2^2 \quad (2)$$

Once an optimal candidate position has been chosen, the process removes it from the distribution and applies again the median minimization criterion of Eq.2 to select another candidate and so on, up to N_{opt}.

3. Single-reference complexity reduction

Up to now, multi-step integration (Sect.2.1) has been presented as an exhaustive motion candidate generation process and previously mentioned computational and memory issues have not been taken into account. In practice, in order to be able to build and to keep in memory the multi-step integration stage outputs, it is necessary to select only a subset of all step

sequences and therefore associated motion paths starting from each pixel $\boldsymbol{x}_{ref}$ of I_{ref}. In what follows, we review the existing algorithmic single-reference strategies which can be considered to perform such computational complexity reduction.

For a given pair of frames $\{I_{ref}, I_n\}$, let N_{max} be the maximum number of motion paths which can be built for each pixel $\boldsymbol{x}_{ref} \in I_{ref}$ according to storage capacity. Limited storage capacity requires the selection of only N_{max} step sequences among the K possible step sequences $\boldsymbol{\Gamma}_{ref,n} = \{\boldsymbol{\gamma}_0, \boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_{K-1}\}$. It leads to $\boldsymbol{\Gamma}^*_{ref,n}$, a set made of N_{max} step sequences selected among $\boldsymbol{\Gamma}_{ref,n}$ with $N_{max} << K$.

3.1. Random selection

The straightforward approach consists in a simple random selection of N_{max} step sequences among $\boldsymbol{\Gamma}_{ref,n}$. However, it induces a systematic bias towards the more populated branches of the tree since the steps assigned to a given intermediate frames between I_{ref} and I_n do not appear with the same frequency. Smallest steps appear more frequently than larger ones among the K possible step sequences and therefore lead to more populated tree branches. In Fig.1a, step 1 assigned to frame I_0 is used in two step sequences ($\{1, 1, 1\}$,$\{1, 2\}$) contrary to steps 2 and 3 which leads to only one step sequence (respectively $\{2, 1\}$ and $\{3\}$). After purely random selection, smallest steps may consequently play a more important role than largest steps among the chosen N_{max} step sequences and it finally may lead to highly correlated resulting candidate displacement vectors.

3.2. Largest step sequences removal

To make the selection more clever, one can remove the largest step sequences in terms of number of constituting steps [18, 19]. In practice, a threshold of K_{max} number of steps can be set and only step sequences $\boldsymbol{\gamma}_i = \{s^i_1, s^i_2, \ldots, s^i_{K_{\boldsymbol{\gamma}_i}}\}$ for which $K_{\boldsymbol{\gamma}_i} \leq K_{max}$ are kept. Indeed, too many steps means too many multi-step optical flow concatenations which may lead to an important motion drift. The goal of removing largest step sequences is more precisely to reduce the effects of the three different error types [21]. First, intrinsic error propagation which deals with accumulation of displacement error along the video sequence. Second, interpolation error which is inherent to the interpolation process since successive motion path positions are non-necessary grid points. Third, motion bias which is bias in motion computation since successive estimated motion path positions are different from the true ones. However, it does not solve the imbalance tree issue due to random selection.

3.3. Step occurrence-based guided random selection

To avoid this issue, uniformizing for all intermediate frames I_c with $ref \leq c \leq n$ the contributions of all steps assigned to I_c is required. In this context, one can constrain the selection using a step occurrence criterion. The idea is to assign a occurrence number to each step of the tree and to update it each time the current step is used in a selected step sequence. The constraint is to tend to make this occurrence of appearance as uniform as possible between all the steps arising from a given frame.

Let $f(s_{cur}, I_c)$ be the occurrence number of step s_{cur} starting from I_c. As described by the pseudo-code in Appendix A, the step occurrence-based guided random selection starts by randomly selecting a first step sequence among $\boldsymbol{\Gamma}_{ref,n}$. The occurrence number of each step of this first step sequence is incremented by 1. Then, each time the algorithm aims at choosing a new step sequence $\boldsymbol{\gamma}_{cur}$, it starts from I_{ref} and iteratively selects among the possible steps $\{s_1, s_2, \ldots, s_{Q_c}\}$ available from the current frame I_c (excepting those for which $c + s_i > n$) the step s_{cur} having the smallest occurrence number, i.e. $f(s_{cur}, I_c) \leq f(s_i, I_c)\ \forall i \in [\![1, 2, \ldots, s_{Q_c}]\!]$ with $s_i \neq s_{cur}$ and $c + s_i \leq n$ and then moves to the subsequent frame $I_{c+s_{cur}}$. If several steps have same occurrence number, a random selection is performed among them. This iterative scheme building $\boldsymbol{\gamma}_{cur}$ is performed until I_n is reached. Before selecting a new step sequence and as done with the first selected one, the occurrence number of each step of $\boldsymbol{\gamma}_{cur}$ is incremented by 1. The whole process is repeated until N_{max} step sequences are chosen.

The limitation wich such guided random selection is that it is too computationally complex to select a different subset of step sequences for each grid point $\boldsymbol{x}_{ref}$ of I_{ref} (i.e. one step sequence tree per pixel $\boldsymbol{x}_{ref}$). Indeed, during motion path construction, it would require to successively load in memory a different set of multi-step optical flow fields once a long-term displacement vector starting from a new $\boldsymbol{x}_{ref}$ is under computation. Thus, the N_{max} selected step sequences of $\boldsymbol{\Gamma}^*_{ref,n}$ are the same for all grid points $\boldsymbol{x}_{ref}$ to allow to build densely and in only one pass all the motion paths starting from all grid points $\boldsymbol{x}_{ref}$.

In practice, each node of the tree obtained through the occurrence-based guided random selection stores an associated dense multi-step optical flow field. Thus, a given node of node value s_i defined from frame I_c stores the field $\boldsymbol{v}_{c,c+s_i}$. Motion paths are built densely by concatenating for each selected step sequences the corresponding multi-step optical flow fields along the tree. For step sequence $\boldsymbol{\gamma}_i \in \boldsymbol{\Gamma}^*_{ref,n}$ for instance, $\boldsymbol{v}_{ref,ref+s^i_1}$ stored into the node of node value s^i_1 and defined at I_{ref} is concatenated with $\boldsymbol{v}_{ref+s^i_1,ref+\sum^2_{m=1} s^i_m}$ and stored in the node corresponding to step s^i_2 starting from $I_{ref+s^i_1}$. The resulting concatenated motion field is then concatenated with $\boldsymbol{v}_{ref+\sum^2_{m=1} s^i_m,ref+\sum^3_{m=1} s^i_m}$ and stored in the node corresponding to s^i_3 starting from $I_{ref+\sum^2_{m=1} s^i_m}$ and so on. By following this dense motion path construction, the leaf nodes of the tree corresponding to step $s^i_{K_{\boldsymbol{\gamma}_i}} \in \boldsymbol{\gamma}_i$ finally store a candidate motion fields $\boldsymbol{d}_{ref,n}$ obtained with the motion path whose corresponding step sequence is $\boldsymbol{\gamma}_i \in \boldsymbol{\Gamma}^*_{ref,n}$. $\boldsymbol{d}_{ref,n}$ will contribute to establish $T_{ref,n}(\boldsymbol{x}_{ref}) = \{\boldsymbol{x}^i_n\}_{i \in [\![0,\ldots,K_{\boldsymbol{x}_{ref}}-1]\!]}$, the set of candidate positions in I_n obtained for each $\boldsymbol{x}_{ref}$ of I_{ref}. By browsing the tree as described above, an optical flow field assigned to a given step used in more than one step sequence of $\boldsymbol{\Gamma}^*_{ref,n}$ is stored and read only one time.

Another limitation is that the intrinsic quality of either optical flows or combinations of optical flows is not taken into account to guide random step sequence selection. A given subset of step sequence may better suit to a given grid point than its neighbor.

To overcome both limitations, we propose to explore an alternative to the step occurrence-based guided random selection which relies on a multi-reference frames processing.

4. MR-MISS: multi-reference frames MISS

We propose a new complexity reduction scheme based on multi-reference frames processing which is relevant for two main reasons. First, it efficiently reduces the complexity during multi-step integration by guiding the step sequence selection using quality criteria which are used to introduce mandatory passage points within the tree of step sequences. Second, it allows to update the appearance of the points under tracking and therefore to make quality criteria more robust to assess the quality of displacement vectors during statistical selection.

As the quality of motion/trajectory fields starting from two given separate areas of I_{ref} temporally decreases differently along the sequence, we cannot anymore compute the same subset of step sequences $\mathbf{\Gamma}^{*}_{ref,n}$ for all grid points of I_{ref}. For this reason, we target the issue of performing a long-term dense motion estimation with respect to a free-form region of interest (ROI) defined in the reference frame I_{ref} and belonging to the same object. Tackling this context is relevant since applications such as video editing tasks often focus on distinct spatial areas. Logo insertion and propagation is a characteristic example.

The proposed algorithm is called multi-reference frames multi-step integration and statistical selection (MR-MISS). It mainly relies on the insertion of new reference frames each time the set of trajectory/motion vectors to be computed start to fail. To reach long-term motion estimation requirements, a MISS strategy (referred to single-reference MISS or SR-MISS in what follows in opposition to MR-MISS) is performed from each inserted reference frames and the resulting multi-reference frames displacement vectors are finally concatenated.

Our approach follows the same spirit of [22] whose aim is to re-correlate short-range sparse pieces of trajectories, called tracklets, estimated with respect to different starting frames in order to go towards longer long-range trajectories. We propose to exploit this concept of tracklets combinations in the context of dense motion estimation.

4.1. MR-MISS overview

We consider a long video shot $\{I_n\}_{n\in[\![0,\ldots,N]\!]}$ of $N+1$ RGB images including the first frame I_0 considered as the main reference frame and denoted as I_{ref_0}. Our goal is to perform long-term dense motion estimation both starting from and with respect to a free-form ROI $\Omega_{ref_0} \in \mathbb{Z}^2$ provided by the user in I_{ref_0}. In this context, we aim at determining with high accuracy and for each pair of frames $\{I_{ref_0}, I_n\}$ with $n \in [\![0,\ldots,N]\!] \neq ref_0$:

- from-the-reference displacement vectors $\boldsymbol{d}_{ref_0,n}$ between pixels $\boldsymbol{x}_{ref_0} \in \Omega_{ref_0}$ in I_{ref_0} and non-necessary grid positions in I_n,
- to-the-reference displacement vectors $\boldsymbol{d}_{n,ref_0}$ from I_n to I_{ref_0}, starting from pixels $\boldsymbol{x}_n \in I_n$ and for which $\boldsymbol{x}_n + \boldsymbol{d}_{n,ref_0}(\boldsymbol{x}_n)$ belongs to Ω_{ref_0}.

These displacement vectors must be accurately computed even if I_n is very distant temporally and if strong content modifications occur between I_{ref_0} and I_n. Instead of relying on a SR-MISS strategy (Sect.3) only referring to the reference frame I_{ref_0}, we suggest to both:

- perform SR-MISS from I_{ref_0} as well as from M intermediate reference frames cleverly inserted within the sequence and referred to I_{ref_k} with $k \in [\![1,2,\ldots,M]\!]$,
- concatenate the resulting multi-reference frames displacement vectors.

With MR-MISS, we give a key role to the quality assessment of trajectory fields. The insertion of the M intermediate reference frames $\{I_{ref_k}\}_{k\in[\![1,2,\ldots,M]\!]}$ is performed automatically by relying on a robust quality assessment of trajectories starting from pixels $\boldsymbol{x}_{ref_0}$ of Ω_{ref_0}. By this way, we add a new intermediate reference frame each time the trajectories under computation diverge. Thus, we continue motion estimation from intermediate sound frames in order to temporally extend the trajectory estimation process with respect to I_{ref_0} as far as possible. Obviously, the intermediate reference frame insertion task can be also performed manually under the operator control, as for specific post-production applications requiring the most realistic viewing experience.

In the following, we describe the MR-MISS strategy by focusing on three aspects: the concatenation of multi-reference frames displacement vectors (Sect.4.2), the reference frames insertion task (Sect.4.3) and the processing of to-the-reference displacement vectors (Sect.4.4). Then, we explain in Sect.4.5 how MR-MISS can be also seen as a robust complexity reduction method in the context of MISS and how it improves long-term dense motion estimation compared to SR-MISS.

4.2. Combination of multi-reference frames vectors

Let us focus on the estimation of the trajectory $\boldsymbol{T}(\boldsymbol{x}_{ref_0})$ starting from the grid point $\boldsymbol{x}_{ref_0} \in \Omega_{ref_0}$ of I_{ref_0}. $\boldsymbol{T}(\boldsymbol{x}_{ref_0})$ is defined by a set of from-the-reference displacement vectors $\{\boldsymbol{d}_{ref_0,n}(\boldsymbol{x}_{ref_0})\}_{n\in[\![1,\ldots,N]\!]}$ which must be accurately estimated for the whole long video shot. Toward this task, we start by applying the SR-MISS algorithm with respect to I_{ref_0}. Let us assume that it fails at I_{fail_0} with $fail_0 < N$ (Fig.2). We propose to introduce a new reference frame at I_{fail_0-1}, i.e. at the instant which precedes the tracking failure and for which $\boldsymbol{d}_{ref_0,fail_0-1}(\boldsymbol{x}_{ref_0})$ has been well estimated.

Once this new reference frame, referred to I_{ref_1}, has been inserted, we run again the SR-MISS algorithm starting from the position $\boldsymbol{x}_{ref_0} + \boldsymbol{d}_{ref_0,ref_1}(\boldsymbol{x}_{ref_0})$ of I_{ref_1} between I_{ref_1} and each subsequent frames I_n with $n \in [\![ref_1+1,\ldots,N]\!]$ (Fig.2). Thus, we obtain the set of displacement vectors $\{\tilde{\boldsymbol{d}}_{ref_1,n}\}_{n\in[\![ref_1+1,\ldots,N]\!]}$ defined with respect to $\boldsymbol{x}_{ref_0} + \boldsymbol{d}_{ref_0,ref_1}(\boldsymbol{x}_{ref_0})$ in I_{ref_1} where $\tilde{.}$ denotes a displacement interpolated at a non-necessary grid position. We can now obtain a new version of the displacement vectors $\{\boldsymbol{d}_{ref_0,n}(\boldsymbol{x}_{ref_0})\}$ with $n \in [\![ref_1+1,\ldots,N]\!]$ by concatenating $\boldsymbol{d}_{ref_0,ref_1}$ estimated with respect to I_{ref_0} and $\tilde{\boldsymbol{d}}_{ref_1,n}$ we just computed with respect to I_{ref_1}:

$$\boldsymbol{d}_{ref_0,n}(\boldsymbol{x}_{ref_0}) = \boldsymbol{d}_{ref_0,ref_1}(\boldsymbol{x}_{ref_0}) + \tilde{\boldsymbol{d}}_{ref_1,n}(\boldsymbol{x}_{ref_0} + \boldsymbol{d}_{ref_0,ref_1}(\boldsymbol{x}_{ref_0})) \quad (3)$$

If this resulting new version of $\boldsymbol{T}(\boldsymbol{x}_{ref_0})$ fails again, at I_{fail_1} for instance, we insert a new reference frame referred to I_{ref_2} at I_{fail_1-1} and we perform SR-MISS with respect to I_{ref_2} (Fig.2).

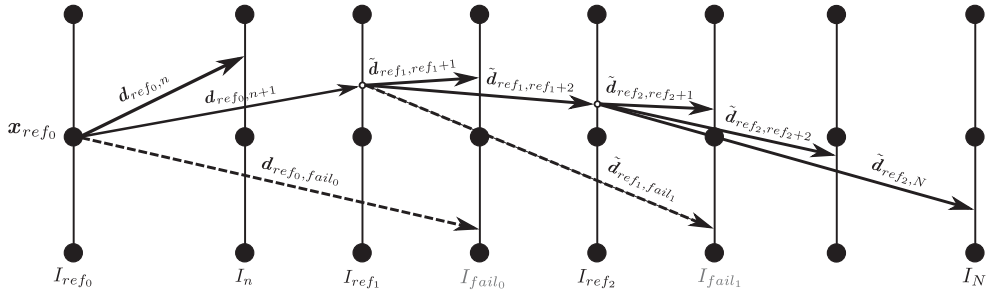


Figure 2: The proposed multi-reference frames MISS strategy (MR-MISS) through insertion of reference frames once trajectories diverge. A robust long-term dense motion estimation is reached by concatenating accurate multi-reference frames displacement vectors (solid vectors) while rejecting motion outliers (dashed vectors).

We finally obtain new estimates of the displacement vectors $\{\boldsymbol{d}_{ref_0,n}(\boldsymbol{x}_{ref_0})\}$ for $n \in [\![ref_2+1,\ldots,N]\!]$:

$$\begin{aligned}\boldsymbol{d}_{ref_0,n}(\boldsymbol{x}_{ref_0}) &= \boldsymbol{d}_{ref_0,ref_1}(\boldsymbol{x}_{ref_0}) + \tilde{\boldsymbol{d}}_{ref_1,ref_2}(\boldsymbol{x}_{ref_0}+\boldsymbol{d}_{ref_0,ref_1}) \\ &+ \tilde{\boldsymbol{d}}_{ref_2,n}(\boldsymbol{x}_{ref_0}+\boldsymbol{d}_{ref_0,ref_1}+\tilde{\boldsymbol{d}}_{ref_1,ref_2}) \qquad (4)\end{aligned}$$

We apply an exactly similar processing each time $\boldsymbol{T}(\boldsymbol{x}_{ref_0})$ fails again, up to the end of the sequence. More generally, by defining I_{ref_q} as the last intermediate reference frame inserted before a given frame I_n, the refined from-the-reference displacement vector between $\boldsymbol{x}_{ref_0}$ of I_{ref_0} and I_n becomes:

$$\boldsymbol{d}_{ref_0,n}(\boldsymbol{x}_{ref_0}) = \boldsymbol{d}_{ref_0,ref_1}(\boldsymbol{x}_{ref_0}) + \sum_{l=1}^{q-1} \tilde{\boldsymbol{d}}_{ref_l,ref_{l+1}}(\boldsymbol{x}_{ref_l}) + \tilde{\boldsymbol{d}}_{ref_q,n}(\boldsymbol{x}_{ref_q}) \qquad (5)$$

where $\boldsymbol{x}_{ref_k}$ defines the successive positions of $\boldsymbol{T}(\boldsymbol{x}_{ref_0})$ in the intermediate reference frames $\{I_{ref_k}\}_{k\in[\![1,2,\ldots,q]\!]}$:

$$\boldsymbol{x}_{ref_k} = \boldsymbol{x}_{ref_0} + \sum_{l=0}^{k-1} \tilde{\boldsymbol{d}}_{ref_l,ref_{l+1}}(\boldsymbol{x}_{ref_l}) \qquad (6)$$

4.3. Selection of intermediate reference frames

We suggest to insert new reference frames based on the detection of tracking failures during the computation of $\boldsymbol{T}(\boldsymbol{x}_{ref_0})$. It can be performed either automatically or interactively through the study of the temporal evolution of matching cost $\mathtt{C}(\boldsymbol{x}_{ref_0}, \boldsymbol{d}_{ref_0,n}(\boldsymbol{x}_{ref_0}))$ and inconsistency $\mathtt{Inc}(\boldsymbol{x}_{ref_0}, \boldsymbol{d}_{ref_0,n}(\boldsymbol{x}_{ref_0}))$ quality features (Appendix B) associated to displacement vectors $\{\boldsymbol{d}_{ref_0,n}(\boldsymbol{x}_{ref_0})\}$ with $n \in [\![1,\ldots,N]\!]$.

In practice, binary matching costs and inconsistency values obtained by thresholding $\mathtt{C}(\boldsymbol{x}_{ref_0}, \boldsymbol{d}_{ref_0,n}(\boldsymbol{x}_{ref_0}))$ and $\mathtt{Inc}(\boldsymbol{x}_{ref_0}, \boldsymbol{d}_{ref_0,n}(\boldsymbol{x}_{ref_0}))$ respectively by $\epsilon_{\mathtt{C}}$ and $\epsilon_{\mathtt{Inc}}$ can inform about the quality of $\boldsymbol{T}(\boldsymbol{x}_{ref_0})$. Once at least one of these thresholds is reached, the current from-the-reference displacement vector is considered as erroneous and the process automatically adds a new reference frame at the previous frame.

To extend the tracking failure detection to the whole set of trajectories starting from pixels $\boldsymbol{x}_{ref_0} \in \Omega_{ref_0}$ in I_{ref_0}, we can focus on the percentage of pixels $\boldsymbol{x}_{ref_0}$ whose corresponding displacement vector is either inaccurate according to binary matching cost or inconsistent according to binary inconsistency value (or both). Thus, we define a threshold $\epsilon_{\%}$ on this percentage to determine from which instants new intermediate reference frames are needed.

Note that with MR-MISS, the ROI must be un-occluded in each intermediate reference frames. Nevertheless, handling temporally occlusions within any temporal section $[I_{ref_k}, I_{ref_{k+1}}]$ is still possible since multi-step optical flows are able to jump between distant frames and therefore to continue the matching process when the entity to be tracked re-appears.

4.4. To-the-reference estimation

If the application under consideration requires the estimation of to-the-reference displacement vectors $\boldsymbol{d}_{n,ref_0}(\boldsymbol{x}_n)$ $\forall n \in [\![1,\ldots,N]\!]$ and with $\boldsymbol{x}_n + \boldsymbol{d}_{n,ref_0}(\boldsymbol{x}_n) \in \Omega_{ref_0}$, as for texture insertion and propagation for instance, we cannot apply the MR-MISS strategy starting from each frame I_n and running back to I_{ref_0} for computational issues. We propose to keep the processing in the from-the-reference direction from I_{ref_0} and therefore to decide the introduction of intermdiate reference frames with respect to the quality of from-the-reference displacement vectors only. A certain correlation between the quality assessment of from-the-reference displacement vectors and the effective quality of to-the-reference displacement vectors is ensured by using inconsistency quality features (Sect.4.3). Inconsistency deals with from/to-the-reference consistency and simultaneously addresses the quality of both vector types. Thus, to-the-reference displacement vectors can benefit from the introduction of these intermediate reference frames anyway. Indeed, unaccurate displacement vectors $\boldsymbol{d}_{n,ref_0}(\boldsymbol{x}_n)$ starting from the grid point $\boldsymbol{x}_n$ of I_n can be refined as follows:

$$\boldsymbol{d}_{n,ref_0}(\boldsymbol{x}_n) = \boldsymbol{d}_{n,ref_q}(\boldsymbol{x}_n) + \sum_{l=1}^{q} \tilde{\boldsymbol{d}}_{ref_{q-l+1},ref_{q-l}}(\boldsymbol{x}_{ref_l}) \qquad (7)$$

where $\boldsymbol{x}_{ref_k}$ defines the successive positions in the intermediate reference frames $\{I_{ref_k}\}_{k\in[\![q,q-1,\ldots,1]\!]}$:

$$\boldsymbol{x}_{ref_k} = \boldsymbol{x}_n + \boldsymbol{d}_{n,ref_q}(\boldsymbol{x}_n) + \sum_{l=k+1}^{q} \tilde{\boldsymbol{d}}_{ref_{q-l+1},ref_{q-l}}(\boldsymbol{x}_{ref_l}) \qquad (8)$$

4.5. Comparison of MR-MISS with SR-MISS

We aim at comparing SR-MISS and MR-MISS strategies by showing both how MR-MISS belongs to the MISS framework

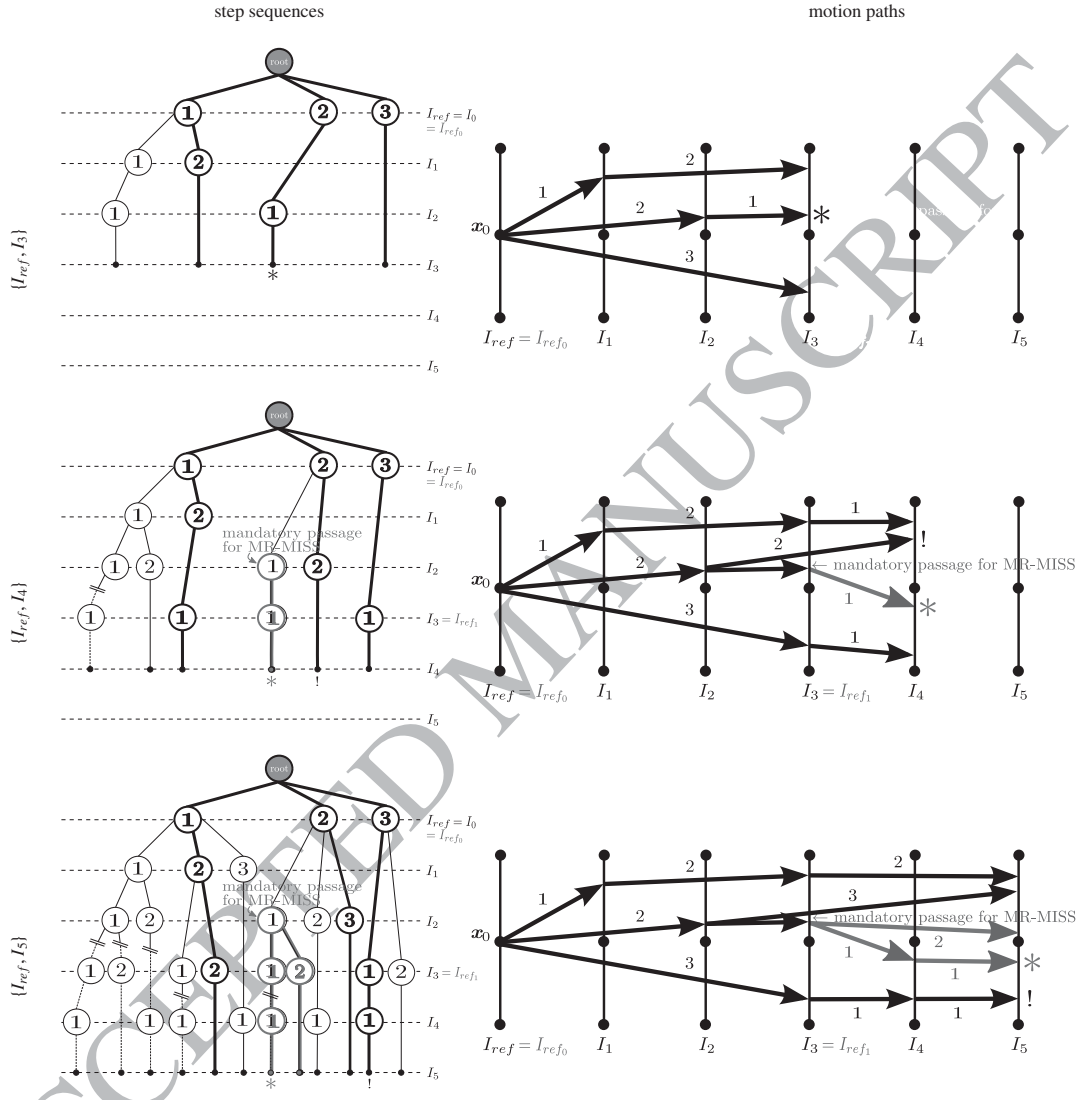


Figure 3: Comparison between SR-MISS and MR-MISS through long-term motion estimation with respect to $\boldsymbol{x}_0 \in I_{ref_0} = I_{ref}$ for the pairs of frames $\{I_{ref}, I_n\}$ with $n = \{3, 4, 5\}$. We assume that only steps 1, 2 and 3 are available from each frame. Moreover, due to computational and memory issues, the maximum number of step sequences to be selected is set to $N_{max} = 3$ and the maximum number of steps within each step sequence is $K_{max} = 3$. Both trees of step sequences and motion paths are presented. Black and grey colors correspond respectively to SR-MISS and MR-MISS. Bold step sequences deal with the $N_{max} = 3$ selected ones. Stars (*) and exclamation marks (!) respectively indicate accurate and unaccurate selected match with respect to ground-truth. Double lines (//) highlight the step sequences which cannot be considered by SR-MISS due to the the maximum concatenation number constraint ($K_{max} = 3$).

and how it reaches better long-term dense motion estimation while efficiently reducing the complexity. In particular, MR-MISS has two main advantages compared to purely random or step occurrence-based random selections (Sect.3).

First, MR-MISS significantly reduces computational complexity since both inserting a reference frame I_{ref_k} with $k \in [\![1, 2, \ldots, M]\!]$ and re-generating motion estimation from I_{ref_k} translates in fixing a mandatory passage point within the tree of step sequences. In terms of motion path generation, it defines non-necessary grid point positions in I_{ref_k} which must necessary belong to the trajectories under estimation and whose accuracy has been checked manually or (semi-)automatically through both cost matching and inconsistency quality features.

Guiding the step selection using quality criteria induces another interesting aspect. Since SR-MISS generates step sequences with a maximum number of steps K_{max} due to computational and memory issues, it does not take in consideration large step sequences whose corresponding motion path would have potentially been accurate. In the context of MR-MISS, applying a SR-MISS strategy from each inserted reference frame I_{ref_k} with $k \in [\![0, 1, \ldots, M]\!]$ allows to consider larger step sequences, i.e. up to $(M + 1).K_{max}$ steps (K_{max} concatenations maximum from each reference frame), regardless computational and memory issues. Indeed, each time a SR-MISS strategy is performed from a new reference frame I_{ref_k}, the motion path generation process has already been done for the temporal segment $[I_{ref_0}, I_{ref_k}]$ and only the subsequent frames undergo the previously mentioned issues.

Second, relying on new reference frames allow to update both the appearance and intermediate positions of grid points $\boldsymbol{x}_{ref_0} \in \Omega_{ref_0}$ in I_{ref_0} under tracking. By this way, both matching cost and inconsistency quality features involved in statistical selection are computed with respect to temporally nearest and accurate intermediate locations. The resulting quality criteria are more valid when we rely on a reference frame which is closer from the current frame than the initial reference frame I_{ref_0}. In particular, our technique apprehends more efficiently smooth changes of appearance of Ω_{ref_0}.

These advantages of MR-MISS compared to SR-MISS are illustrated in Fig.3. In this example, we focus on long-term motion estimation with respect to $\boldsymbol{x}_0 \in I_{ref_0} = I_{ref}$ and we illustrate how SR/MR-MISS work for the pairs of frames $\{I_{ref}, I_n\}$ with $n = \{3, 4, 5\}$. We assume that only steps 1, 2 and 3 are available from each frame. Moreover, due to computational and memory issues, the maximum number of step sequences to be selected and the maximum number of steps within each step sequence are respectively set to $N_{max} = 3$ and $K_{max} = 3$.

Among the 4 step sequences available between I_{ref} and I_3, N_{max} of them are selected ($\{\{1, 2\}, \{2, 1\}, \{3\}\}$) according to the step occurrence-based guided random selection. Finally, the motion path corresponding to $\{2, 1\}$ leads to the best match in I_3 and this matching appears as accurate.

For SR-MISS, the processing for $\{I_{ref}, I_4\}$ is performed independently from the one of $\{I_{ref}, I_3\}$ and leads to the selection of $\{\{1, 2, 1\}, \{2, 2\}, \{3, 1\}\}$. Among the 3 resulting motion paths, the one built via the step sequence $\{2, 2\}$ is selected but leads to an unaccurate matching both according to quality criteria and ground-truth (true detection of motion outlier) without being able to propose a more efficient alternative. This tracking failure in I_4 forces MR-MISS to select I_3 as a new intermediate reference frame referred to I_{ref_1}. Thus, MR-MISS matches $\boldsymbol{x}_0$ to an accurate position in I_4 by concatenating the inter-reference displacement vector $\boldsymbol{d}_{ref_0,ref_1}$ and the optical flow of step 1 between I_{ref_1} and I_4 as in Eq.3. By starting from a much closer sound reference frame and therefore by forcing the selected step sequence to start by the sub-sequence $\{\mathbf{2,1}\}$, MR-MISS succeeds in finding the optimal step sequence ($\{\mathbf{2,1}, 1\}$) while limiting the computational complexity. In this simple example, only one step sequence is considered by MR-MISS between I_{ref} and I_4 whereas SR-MISS has to make a selection among $\{\{1, 1, 2\}, \{1, 2, 1\}, \{2, 1, 1\}, \{2, 2\}, \{3, 1\}\}$.

Let us go further by studying how SR/MR-MISS process $\{I_{ref}, I_5\}$. Contrary to SR-MISS which reaches another bad match by selecting the motion path corresponding to $\{3, 1, 1\}$, MR-MISS remains less computationally complex by allowing the selection only between motion paths of step sequences $\{\mathbf{2,1}, 1, 1\}$ and $\{\mathbf{2,1}, 2\}$. $\{2, 1, 1, 1\}$ is supposed to give the optimal match in our example. Another interesting aspect is that many step sequences, such as the optimal one $\{2, 1, 1, 1\}$, cannot be considered by SR-MISS due to the limitation in terms of number of concatenations ($K_{max} = 3$) since $\{2, 1, 1, 1\}$ is made of 4 steps. We illustrate by this way how MR-MISS reduces the computational complexity while tending to keep in the reduced step sequence set the more optimal ones.

5. Experimental results

5.1. Complexity reduction and accuracy improvement

Complexity reduction and accuracy improvements are the two expected points when comparing MR-MISS to SR-MISS. Additionally to the analytical comparison performed in Sect.4.5, we provide here the results of a comparative experimental assessment performed on the pair $\{I_{ref} = I_0, I_{25}\}$ of the *Walking Couple* sequence. Its goal translates in evaluating via Peak Signal to Noise Ratio (PSNR) the quality of the warping from I_0 to I_{25} of a free-form ROI of size 70×160 pixels corresponding to the yellow shape inserted in Fig.5.

ROI warping from I_0 to I_{25} is performed using from-the-reference displacement vectors estimated either with SR-MISS or MR-MISS and this for a maximum number of possible motion paths N_{max} varying from 6 to 60 in order to simulate different limited computational and storage capacity characteristics. The range used for N_{max} is of the same order of magnitude as one can encounter with standard computers (< 100). The maximum number of concatenations and the number of candidate positions selected by statistical processing are respectively $K_{max} = 7$ and $N_{opt} = 3$. In this experiment, we use Large Displacement Optical Flow (LDOF) optical flows [23] previously estimated with the following set of empirically selected steps: $\{1, 2, 3, 4, 5, 10, 15\}$.

SR-MISS processes the sequence from $\{I_{ref}, I_1\}$, $\{I_{ref}, I_2\}$ ... to $\{I_{ref}, I_{25}\}$ with the different selected N_{max} values. For each one, SR-MISS is run 7 times. These multiple realizations aim

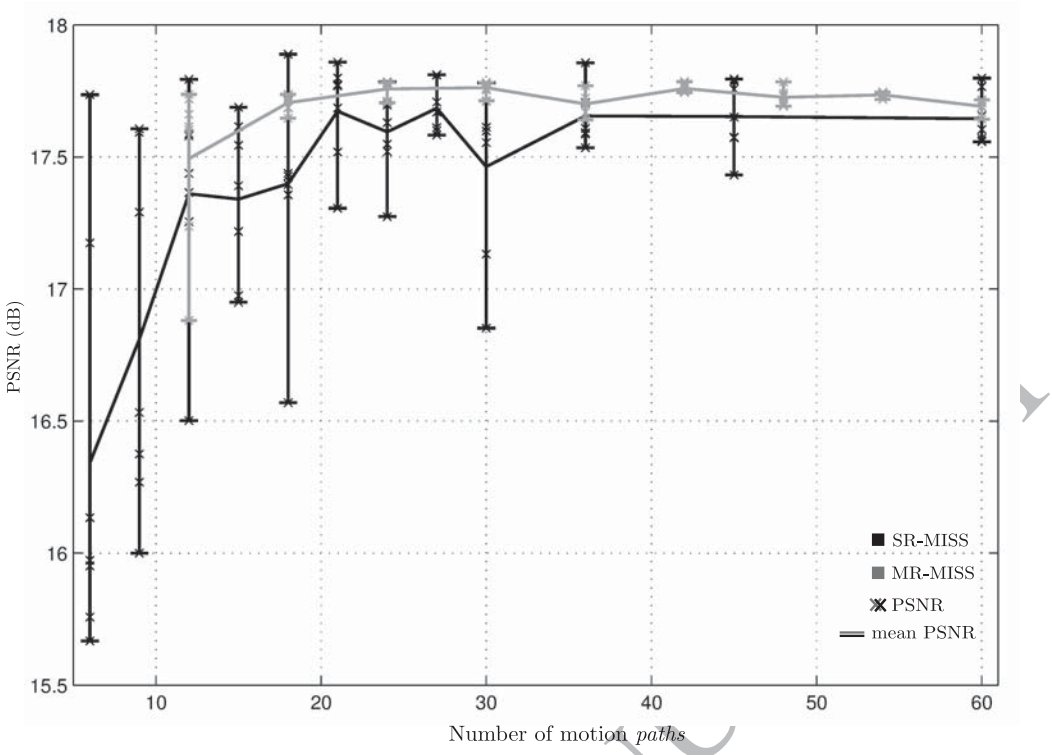


Figure 4: Assessment of the warping quality induces by from-the-reference displacement vectors estimated either via SR-MISS or MR-MISS for a free-form ROI (yellow shape inserted in Fig.5) between I_0 and I_{25} of the *Walking Couple* sequence. The warping quality is studied via PSNR for a varying number of motion paths simulating limited computational and storage capacity. Black and grey colors correspond to SR-MISS and MR-MISS respectively. Crosses indicate PSNR for each realization whereas curves focus on mean PSNR over the 7 realizations performed for each motion paths number. Bars are drawn from worse to best PSNR values in order to highlight the warping quality variability.

at obtaining an average warping quality since one single processing gives variable results due to MISS random aspects. An intermediate reference frame is then manually set at I_{18} in order to make MR-MISS runs with I_0 and I_{18} as reference frames. The temporal section $[I_{19}, I_{25}]$ is processed with the same computational and storage capacity as $[I_0, I_{18}]$ (i.e. same N_{max} from each reference frame $\{I_0, I_{18}\}$). Moreover, for a given N_{max} value, MR-MISS relies on the SR-MISS realization of worse PSNR among the 7 realizations performed with same N_{max} for the temporal section $[I_0, I_{18}]$. By this way, we do not favor MR-MISS with respect to SR-MISS.

Fig.4 displays all the PSNR obtained with respect to the number of motion paths involved between I_0 and I_{25}: N_{max} for SR-MISS, $2\times N_{max}$ for MR-MISS since MR-MISS involves 2 reference frames. Thus, we assess the performance of both methods while setting similar complexity characteristics.

Two findings arise when analysing Fig.4. First, whatever the method, higher the number of motion paths used, the higher the PSNR. Increasing the complexity also results in a decrease of the PSNR standard deviation since the probability of selecting an optimal motion path is obviously correlated to the size of the motion path set. Second, at equal number of motion paths, motion estimation is more accurate with MR-MISS than SR-MISS according to mean PSNR values. With 30 involved motion paths, the gain in terms of PSNR is about 0.3 (17.76 against 17.46). Moreover, the warping quality variability is less important with MR-MISS since it tends to keep more accurate motion paths than SR-MISS and therefore to decrease the probability of tracking failures.

5.2. Benefits of updating ROI appearance

Updating the appearance of the area under tracking at each new intermediate reference frame is a key aspect which explains the good performance of MR-MISS. To quantitatively assess the gain reached by this appearance update, MR-MISS has been applied on the *Walking Couple* sequence (Fig.5) with 2 different presets: the original algorithm described in Sect.4 and MR-MISS without appearance update which means that during statistical selection, matching cost assigned to each candidate displacement vector refers to the color appearance in I_{ref_0} instead of I_{ref_q}, the lastly inserted intermediate reference frame. This latter aspect is directly linked to the SR-MISS strategy which always relies on the initial reference frame I_{ref_0} to compute the displacement vector quality.

The experiment focuses on the smooth changes of color appearance of the free-form ROI involved in the assessment of Sect.5.1. As previously, the reference frames are $\{I_0, I_{18}\}$, whatever the MR-MISS preset. The processing for the temporal section $[I_0, I_{18}]$ is done to provide similar results. In the preset without appearance update, the reference frame I_{18} is replaced by a warped version of I_0 built using the displacement vectors computed for the pair $\{I_0, I_{18}\}$. The MISS procedure with respect to I_{18} is then performed as if the ROI appearance had not

MR-MISS	$\{I_0, I_{19}\}$	$\{I_0, I_{30}\}$	$\{I_0, I_{35}\}$	$\{I_0, I_{40}\}$	$\{I_0, I_{41}\}$	$\{I_0, I_{42}\}$
without AD	17.78	15.37	15.24	16.25	16.12	15.47
with AD	**17.82**	**15.47**	**15.51**	**16.40**	**16.45**	**15.72**

Table 1: Assessment of the gain reached by the MR-MISS appearance update (AD) via mean PSNR scores averaged over 7 realizations for several pairs from $\{I_0, I_{19}\}$ to $\{I_0, I_{42}\}$ of the *Walking Couple* sequence (ROI corresponding to the yellow shape inserted in Fig.5). Best results are in bold.

undergone any changes contrary to the original preset which updates the content appearance in I_{18}.

To compare from-the-reference displacement vectors coming from both presets, Tab.1 shows mean PSNR scores averaged over 7 realizations for several pairs from $\{I_0, I_{19}\}$ to $\{I_0, I_{42}\}$. The PSNR is better for all pairs with appearance update in I_{18}. In addition, the trend is that the PSNR gain rises up with the temporal distance to I_0 which indicates that motion drift is delayed temporally thank to the ROI appearance update. This main MR-MISS aspect tends to achieve long-term requirements even with smooth color or illumination variations along the sequence.

5.3. Video editing

Three video editing examples are provided to qualitatively assess the good performance of MR-MISS in comparison to SR-MISS for different types of complex scenes. The experiment consists in propagating a logo/texture inserted in I_{ref} across the sequence using SR/MR-MISS to-the-reference displacement vectors to bring back inserted data into each current frame. Three sequences are considered: *Walking Couple* between I_0 and I_{60} (Fig.5), *MPI-S1* between I_{115} and I_{165} (Fig.6) and *Hope* between I_{5036} and I_{5111} (Fig.7).

The thresholds related to tracking failure detection are set to $\epsilon_{\mathtt{C}} = 3$, $\epsilon_{\mathtt{Inc}} = 1$ and $\epsilon_{\%} = 0.5$. In addition, $N_{max} = 90$, $K_{max} = 7$ and $N_{opt} = 3$. The input optical flow estimators and the set of steps are specified for each experiment. Small steps from 1 to 5 are used in any case, as one expects. For larger steps, it depends on possible temporary occlusions which may occur in the sequence. At least one step bigger than the duration of the temporary occlusion is required to be able to jump it.

In terms of computation time, performing MR-MISS on a sequence of 700x700 frames (as for *Walking Couple*) takes approximately about 6 minutes per frame on a laptop PC equipped with a 2,5GHz Intel Core i5 processor.

Walking Couple. SR/MR-MISS to-the-reference displacement vectors $\boldsymbol{d}_{n,0}\ \forall n \in [\![1, \ldots, 60]\!]$ are used to propagate the yellow texture across the sequence, from I_0 to I_{60} (Fig.5). The texture is inserted within the shirt of the woman exhibiting periodic structures, highly non-rigid motion as well as illumination changes. Both methods use LDOF [23] optical flows with steps $\{1, 2, 3, 4, 5, 10, 15\}$.

The two first rows and the fourth one show how the SR-MISS motion estimation from I_0 performs the propagation. We notice that a hole appears in I_{31} (upper right part of the texture) and grows gradually due to bad motion estimation of the periodic structures. It also appears that the compacity of the initial texture is lost from in I_{48}. The texture diverges abnormally above and to the right of the correct texture position.

The third and the fifth rows illustrate how MR-MISS performed with respect to reference frames $\{I_0, I_{27}, I_{42}\}$ propagates the texture up to I_{60}. Despite small holes, the results appear to be much better than the ones obtained with SR-MISS. The propagation is clearly performed without any disturbing artefacts. We can also notice that the occlusion due to the arm of the woman is well handled by our method. Occluded parts of the texture are not propagated, as one expects.

MPI-S1. A logo is inserted in an un-textured area which undergoes strong illumination variations as well as a non-rigid transformation due to the rotation of the woman (Fig.6). Additionally to I_{115}, three intermediate reference frames ($\{I_{135}, I_{155}, I_{160}\}$) are inserted for MR-MISS. 2D-DE [24] optical flows as used as inputs with steps $\{1, 2, 3, 4, 5, 8, 10, 15, 20\}$.

The first row shows good results for the 16 first frames. Then, by comparing the second and the third row, we notice that SR-MISS makes the logo progressively distorted from I_{139} and finally not at all recognizable (I_{151}). On the contrary, MR-MISS keeps the logo in a compact form and accurately follows the non-rigid motion of the woman. Finally, the fourth row indicates that it is possible with MR-MISS to rely on good motion estimates for a temporal distance of 50 frames.

Hope. Fig.7 shows logo insertion in a uniform area of I_{5036} and propagation along the *Hope* sequence up to I_{5111}. SR-MISS from I_{5036} and MR-MISS with $\{I_{5036}, I_{5063}, I_{5073}\}$ as reference frames are performed using 2D-DE optical flows with steps $\{1, 2, 3, 4, 5, 8, 10, 15, 20\}$. We notice that SR-MISS makes holes appear and highly distorts the initial logo shape from I_{5103}. On the contrary, visual results with MR-MISS reveal a better consistency over time, up to I_{5111}.

The two last video editing experiments reveal good accuracy with the 2D-DE [24] optical flow which has not been designed especially to be more robust to large movements than common optical flow algorithms, contrary to LDOF [23]. The performance reached by MR-MISS is obviously related to the input optical flow estimator. However, results suggest that MR-MISS is robust enough to extend the ability of its input estimator to go towards longer long-term dense motion estimation.

5.4. Comparison with ground-truth

Quantitative results have been obtained using dense ground-truth trajectory data provided by the *Flag* benchmark dataset [15]. This dataset is based on sparse motion capture data estimated on a flag waving in the wind. Sparse estimates have been interpolated to create a dense 3D surface which has been then projected into the image plane to provide dense ground-truth data. The original version of the resulting *Flag* sequence, displayed in Fig.8, has been used to test MR-MISS, SR-MISS as well as state-of-the-art methods. Experiments focus on direct motion estimation between each pair $\{I_{ref}, I_n\}$ using LDOF [23] (LDOF direct), ITV-L1 [26] (ITV-L1 direct) and

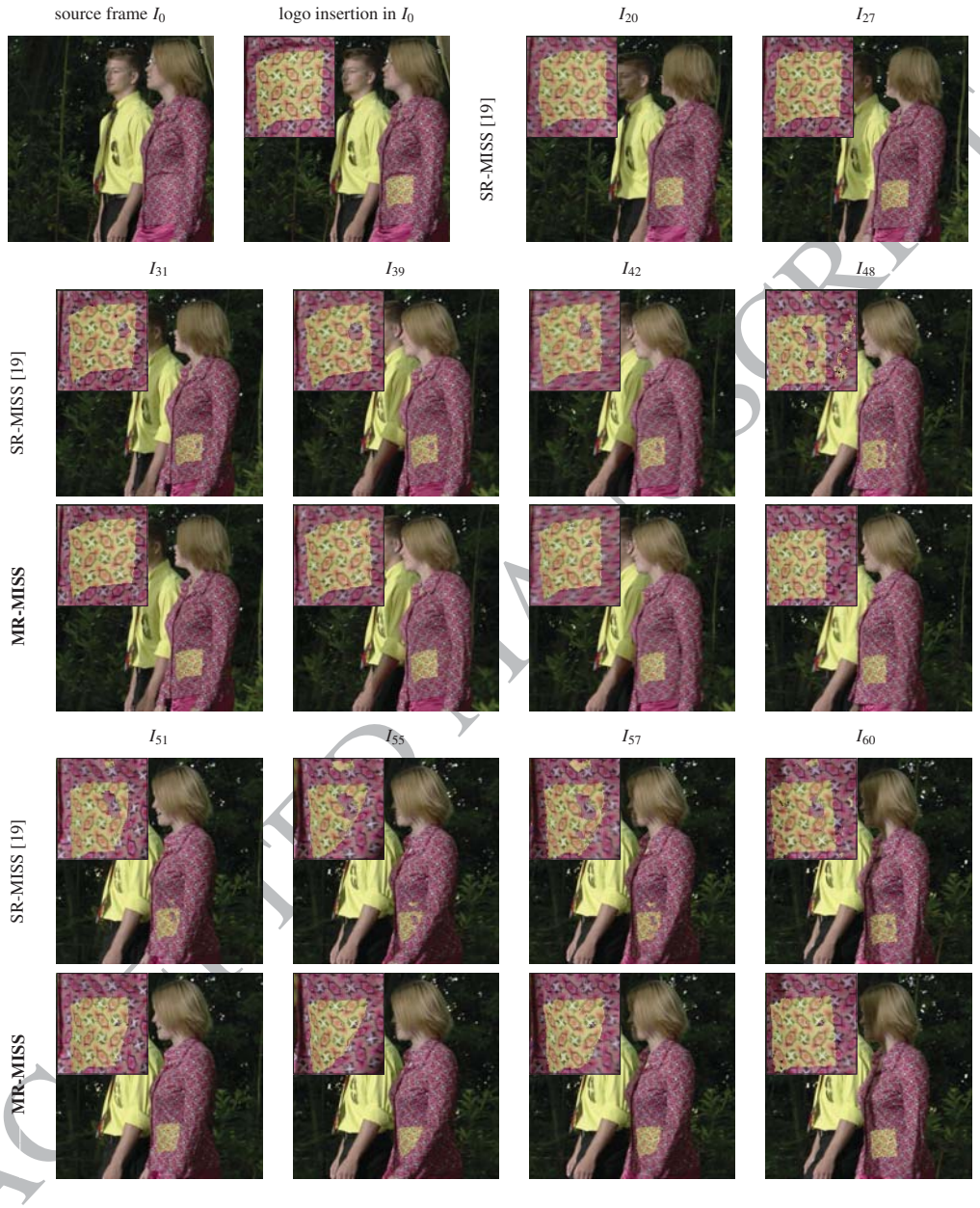


Figure 5: Texture insertion in I_0 and propagation along the *Walking Couple* sequence up to I_{60}. We compare SR-MISS [19] and the proposed MR-MISS strategy using $\{I_0, I_{27}, I_{42}\}$ as reference frames. Both methods use LDOF input optical flows from [23].

Figure 6: Texture insertion in I_{115} and propagation along the *MPI-S1* sequence up to I_{165}. We compare SR-MISS [19] and the proposed MR-MISS strategy using $\{I_{115}, I_{135}, I_{155}, I_{160}\}$ as reference frames. Both methods use 2D-DE input optical flows from [24].

Figure 7: Texture insertion in I_{5036} and propagation along the *Hope* sequence up to I_{5111}. We compare SR-MISS [19] and the proposed MR-MISS strategy using $\{I_{5036}, I_{5063}, I_{5073}\}$ as reference frames. Both methods use 2D-DE input optical flows from [24].

Figure 8: Source frames of the *Flag* sequence [25].

the keypoint-based non-rigid registration algorithm described in [27] ([27] direct), classical Euler integration via concatenation of LDOF optical flows computed between consecutive frames (LDOF acc), multi-frame subspace flow (MFSF) proposed in [25] and its extended version detailed in [15] using PCA or DCT trajectory basis (MFSF-PCA, MFSF-DCT) as well as MSF [17], SR-MISS and MR-MISS with LDOF optical flows of steps $\{1, 2, 3, 4, 5, 8, 10, 15, 20, 25, 30, 40, 50\}$. MR-MISS uses $\{I_0, I_{20}\}$ as reference frames.

All these methods are compared through Root Mean Square (RMS) endpoint errors between the respective obtained from-the-reference displacement fields and the ground-truth data (Tab.2). The RMS errors are estimated for all the foreground pixels and for all the pairs of frames $\{I_{ref}, I_n\}$ together. RMS endpoint errors computed for each pair of frames are also shown in Fig.9 for all the methods based on LDOF: LDOF direct, LDOF acc, MSF, SR-MISS and MR-MISS.

In Tab.2, we notice that MR-MISS outperforms SR-MISS with a global RMS error of 0.58 pixels against 0.69. A for-

Method	RMS error
MR-MISS	**0.58**
SR-MISS [19]	0.69
MSF [16, 17]	1.41
LDOF direct [23]	1.74
LDOF acc [23]	4
MFSF-PCA [15]	0.69
MFSF-DCT [15]	0.80
MFSF-PCA [25]	0.98
MFSF-DCT [25]	1.06
Pizarro et al. [27] direct	1.24
ITV-L1 direct [26]	1.43

Table 2: RMS endpoint errors (in pixel) for different methods on the *Flag* benchmark dataset [25]. The best result is in bold.

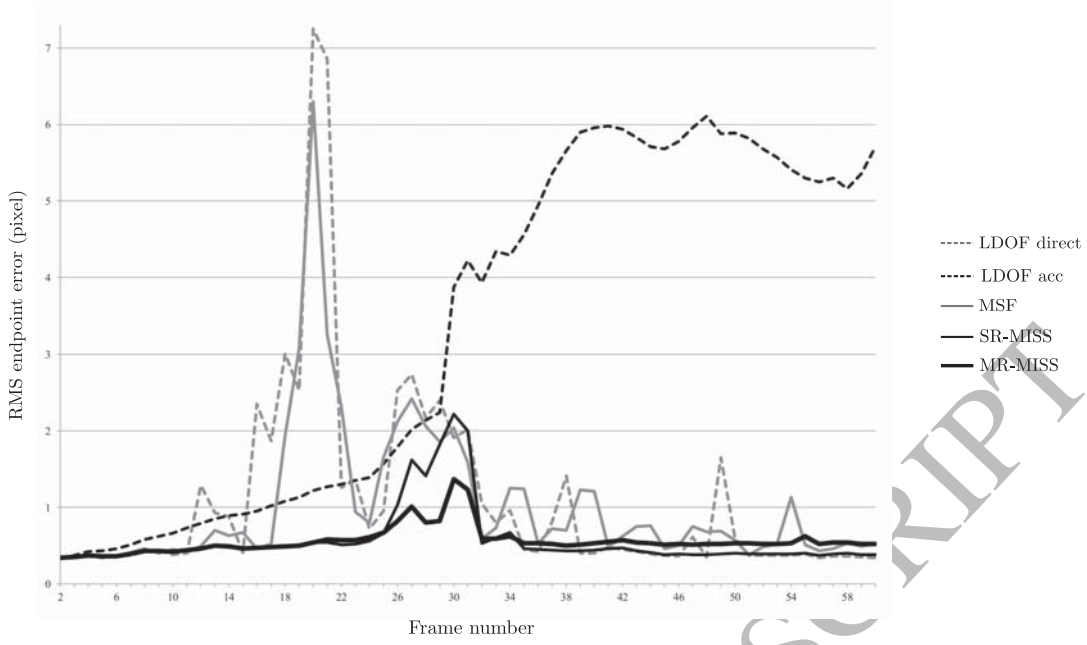


Figure 9: RMS endpoint errors for each pair $\{I_{ref}, I_n\}$ along *Flag* sequence [25] with LDOF direct, LDOF acc, MSF [16, 17], SR-MISS [19] and our multi-reference frames strategy MR-MISS with $\{I_1, I_{20}\}$ as reference frames.

tiori, it gives more accurate displacement fields than all the single reference frame methods including the challenging MFSF-PCA [15]. When studying the RMS endpoint errors computed for each pair of frames with LDOF direct, LDOF acc, MSF, SR-MISS and MR-MISS (Fig.9), we observe that MR-MISS shows a clear improvement compared to LDOF direct, LDOF acc and MSF. In addition, compared to SR-MISS, MR-MISS strongly reduces the matching issues around I_{30} which coincides with the maximum deformation of the flag (Fig.8). Indeed, SR-MISS gives a RMS error of 2.22 pixels for the pair $\{I_0, I_{29}\}$ whereas MR-MISS leads to a RMS error of 1.37 pixels.

However, MR-MISS gives slightly worse results from I_{35} to I_{60} due to the fact that the flag comes back approximately to its initial position at the end of the sequence (Fig.8). In this very particular context of an almost symmetric sequence, it appears that the matching criteria are more valid with respect to I_0 than with respect to I_{20}. More generally, starting from the same optical flow estimator, Fig.9 proves that MR-MISS is competitive compared to challenging state-of-the-art methods.

The same findings arise when comparing LDOF direct, LDOF acc, MSF, SR-MISS and MR-MISS applied on the *Head* sequence, one of the longest sequences taking part of the *Hopkins-155* dataset [28]. Made of 60 frames, *Head* (Fig.11) is provided with ground-truth trajectories starting from a sparse set of 99 pixels of I_1. All those pixels are visible along the whole sequence. MSF, SR-MISS and MR-MISS are performed using LDOF optical flow with steps $\{1, 2, 3, 4, 5, 10, 20\}$. MSF and SR-MISS use I_1 as reference frame whereas MR-MISS focuses on $\{I_1, I_{35}\}$. The comparisons between the ground-truth and the estimated trajectories displayed in Fig.10 involve two error measures: the position median absolute error (MAE) as well as the percentage of points of I_1 whose location in I_n with $n \in [\![2, \ldots, 60]\!]$ is distant of maximum 1 pixel with respect to the ground-truth locations. The comparative study reveals globally better results for MISS algorithms, especially compared to LDOF acc and MSF. In particular, we notice the ability of both SR-MISS and MR-MISS to reach a longer long-term estimation with MAEs under 1 pixel after 60 frames (0.76 for MR-MISS against 1.60 for MSF). Fig.10 also highlights a slightly better accuracy using MR-MISS with respect to SR-MISS. Thus, MAE decreases from 0.76 to 0.61 pixel in I_{59} (Fig.10a). The gain in terms of percentage of erroneous positions is 8.9% in I_{54} (Fig.10b).

Finally, to illustrate the latter results, a video editing example on the *Head* sequence is provided in Fig.11. A red circular texture is inserted into a uniform area of I_1 and propagated up to I_{60} using LDOF direct, LDOF acc, MSF, SR-MISS and MR-MISS. The texture propagated with LDOF direct and MSF knows significant distortions, as denoted in I_{45} and I_{60}. The circular shape is not maintained due to strong rotation of the head and we observe in both cases a large drift from the cheek to the eye of the character. Reduced drift and distortions remain for LDOF acc. Conversely, both texture compactness and positions are respected through SR-MISS and MR-MISS whose results are clearly better. Compared to SR-MISS, slight improvements can be perceived with MR-MISS, especially in I_{45}.

6. Conclusion

In this work, we addressed long-term dense motion estimation via multi-step integration and statistical selection (MISS). In a combinatorial fashion, MISS-based methods combine optical flows estimated with various inter-frame distances to reach a good compromise between consecutive optical flow concatena-

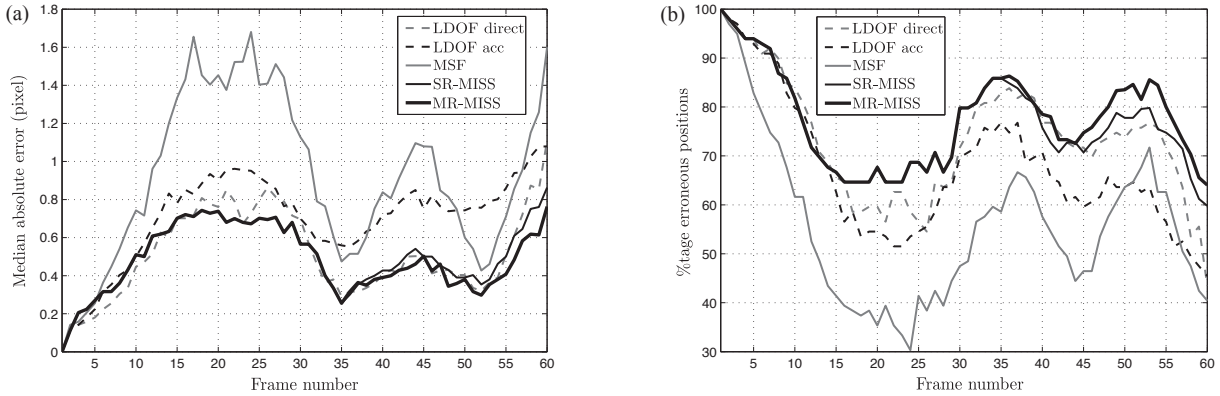


Figure 10: Position median absolute errors (a) and percentage of points of I_1 whose location in I_n is distant of maximum 1 pixel with respect to the ground-truth locations (b) along the *Head* sequence [28] with LDOF direct, LDOF acc, MSF [16, 17], SR-MISS [19] and our multi-reference frames strategy MR-MISS with $\{I_1, I_{35}\}$ as reference frames.

tions which is prone to motion drift and direct matching which is sensitive to ambiguous correspondences.

Managing numerous combinations of multi-step optical flows towards robust statistical selection requires a complexity reduction scheme to overcome computational and memory issues. The main contribution of this work deals with a new strategy, called MR-MISS, which reaches long-term requirements more accurately than existing methods while efficiently reducing the complexity. Based on the re-correlation of dense multi-reference frames trajectories following robust quality criteria such as matching cost and inconsistency, the proposed approach is perfectly suited for challenging applications like video editing tasks. In practice, MR-MISS inserts intermediate reference frames each time motion estimation fails and performs again multi-step integration and statistical selection from these intermediate sound frames.

By this way, MR-MISS guides the step selection by fixing mandatory passage within the tree of step sequences. It reduces the complexity while keeping in the reduced set of resulting motion paths the more accurate ones. With same computational and memory constraints, it allows to consider larger step sequences than existing MISS schemes. Another key aspect is the appearance update which occurs from each inserted intermediate reference frame. It delays motion drift while making quality criteria more valid in case of smooth appearance variations.

Compared to state-of-the-art, MR-MISS significantly improves dense from-the-reference and to-the-reference displacement vectors quality over extended periods of time. It is especially true for complex sequences featuring periodic structures, poorly textured areas and highly non-rigid motion. In this context, significant improvements and good intrinsic performance have been shown quantitatively through warping quality scores and comparisons to dense ground-truth data as well as qualitatively using texture and logo propagation.

Other aspects must deserve more attention for further research. First, MR-MISS has been presented in the context of long-term dense motion estimation with respect to a free-form region of interest towards robust video editing. Its extension to the whole image could be easily reached by performing such strategy to a set of super-pixels covering the whole scene. Moreover, it could be judicious to inject into the set of candidate correspondences the outputs of rigid and deformable motion models as well as sparse KLT of SIFT correspondences. Color variations could also be more explicitly modeled by either introducing color or luminance gain factors while computing displacement vector quality or regularizing in terms of gain similarities. Finally, estimating displacements by relying on multiple reference frames is a first step towards a complete coverage of all the visible points in all the frames. How to compactly represent the totality of the video content with associated long-term motion behavior while taking into account local and global variations of illuminations is however an open challenge.

Appendixes

A. We present the pseudo-code for the step occurrence-based guided random selection described in Section 3.3 for the pair of frames $\{I_{ref}, I_n\}$. As inputs, it takes $\mathbf{\Gamma}_{ref,n} = \{\boldsymbol{\gamma}_0, \boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_{K-1}\}$ to obtain as outputs $\mathbf{\Gamma}^*_{ref,n}$, the set made of N_{max} motion paths selected among $\mathbf{\Gamma}_{ref,n}$.

begin
  $\mathbf{\Gamma}^*_{ref,n} \leftarrow \{\}$
  for $c = ref$ to $n-1$ **do**
    for $l = 1$ to Q_c **do**
      if $c + s_l \leq n$ **then**
        $f(s_l, I_c) \leftarrow 0$ {*step* frequency initialization}
      end if
    end for
  end for
  $\boldsymbol{\gamma}_{cur} \leftarrow \{\}$ {current motion path initialization}
  for $k = 1$ to N_{max} **do**
    if $k = 1$ **then**
      $\boldsymbol{\gamma}_{cur} \leftarrow$ random selection among $\mathbf{\Gamma}_{ref,n}$
    else
      $c = ref$ {current frame number initialization}
      while $I_c < I_n$ **do**

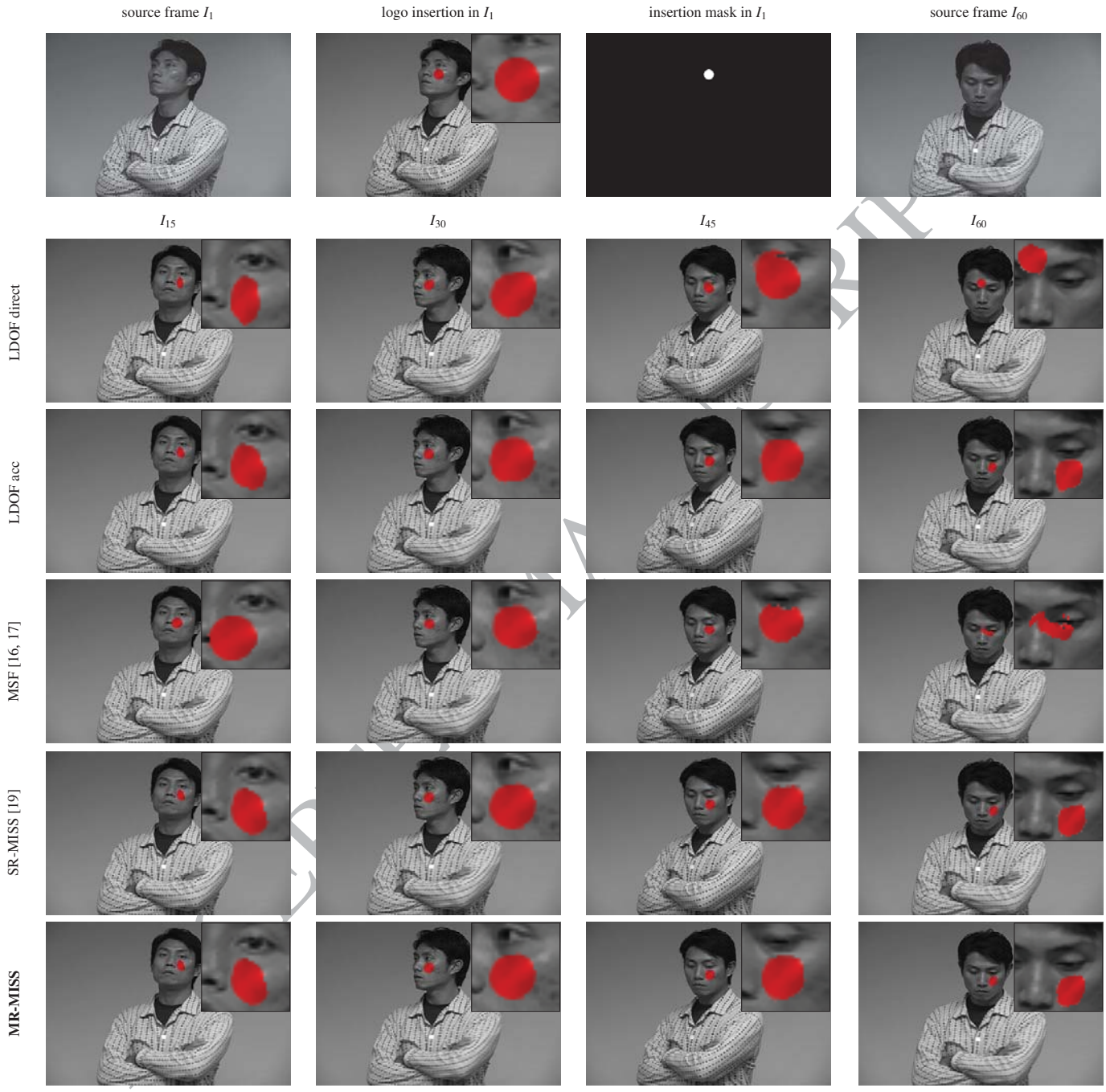


Figure 11: Texture insertion in I_1 and propagation along the *Head* sequence [28] up to I_{60}. We compare LDOF direct, LDOF acc, MSF [16, 17], SR-MISS [19] and the proposed MR-MISS strategy using $\{I_1, I_{35}\}$ as reference frames. All these methods use LDOF input optical flows from [23].

$s_{cur} \leftarrow arg \min_{s_l \in \{s_1, s_2, \ldots, s_{Q_c}\} | c+s_l \leq n} f(s_l, I_c)$
$c \leftarrow c + s_{cur}$
$\boldsymbol{\gamma}_{cur} \leftarrow \{\boldsymbol{\gamma}_{cur}, s_{cur}\}$
end while
end if
$\boldsymbol{\Gamma}^*_{ref,n} \leftarrow \{\boldsymbol{\Gamma}^*_{ref,n}, \boldsymbol{\gamma}_{cur}\}$
$c = ref$ {current frame number initialization}
for $l = 0$ to $K_{\boldsymbol{\gamma}_{cur}} - 1$ **do**
$f(s_l, I_c) \leftarrow f(s_l, I_c) + 1$
$c \leftarrow c + s_l$
end for
end for
end

B. We provide a brief description of both matching cost and inconsistency quality features, involved for statistical selection (Sect.2.2) and tracking failure detection (Sect.4.3).

Matching cost. To indicate how accurately a pixel of I_{ref_0} can be reconstructed by its matched point in I_n, the matching cost $\mathtt{C}(\boldsymbol{x}_{ref_0}, \boldsymbol{d}_{ref_0,n}(\boldsymbol{x}_{ref_0}))$ computes the absolute difference between the RGB color values of $\boldsymbol{x}_{ref_0}$ with those of $\boldsymbol{x}_{ref_0} + \boldsymbol{d}_{ref_0,n}(\boldsymbol{x}_{ref_0})$ in I_n, as written below:

$$\mathtt{C}(\boldsymbol{x}_a, \boldsymbol{d}_{a,b}(\boldsymbol{x}_a)) = \sum_{c \in \{r,g,b\}} I^c_a(\boldsymbol{x}_a) - \tilde{I}^c_b(\boldsymbol{x}_a + \boldsymbol{d}_{a,b}(\boldsymbol{x}_a))$$

Inconsistency. The inconsistency quality feature relies on both from/to-the-reference displacement vectors estimated between I_{ref_0} and I_n. The inconsistency value $\mathtt{Inc}(\boldsymbol{x}_{ref_0}, \boldsymbol{d}_{ref_0,n}(\boldsymbol{x}_{ref_0}))$ associates an intrinsic continuous quality value to $\boldsymbol{d}_{ref_0,n}(\boldsymbol{x}_{ref_0})$ by assessing the consistency between the from-the-reference displacement vector $\boldsymbol{d}_{ref_0,n}(\boldsymbol{x}_{ref_0})$ starting from $\boldsymbol{x}_{ref_0} \in I_{ref_0}$ and its corresponding to-the-reference displacement vector running from I_n to I_{ref_0} and starting from $\boldsymbol{x}_{ref_0} + \boldsymbol{d}_{ref_0,n}(\boldsymbol{x}_{ref_0})$:

$$\mathtt{Inc}(\boldsymbol{x}_a, \boldsymbol{d}_{a,b}(\boldsymbol{x}_a)) = \| \boldsymbol{d}_{a,b}(\boldsymbol{x}_a) + \tilde{\boldsymbol{d}}_{b,a}(\boldsymbol{x}_a + \boldsymbol{d}_{a,b}(\boldsymbol{x}_a)) \|_2$$

References

[1] B. Lucas, T. Kanade, An iterative image registration technique with an application to stereo vision, International joint conference on artificial intelligence 2 (1981) 674–679.

[2] B. Horn, B. Schunck, Determining optical flow, Artificial Intelligence 17 (1) (1981) 185–203.

[3] P. Pérez, C. Hue, J. Vermaak, M. Gangnet, Color-based probabilistic tracking, in: European Conference on Computer Vision, 2002, pp. 661–675.

[4] J. Shi, C. Tomasi, Good features to track, in: IEEE International Conference on Computer Vision and Pattern Recognition, 1994, pp. 593–600.

[5] T. Brox, J. Malik, Object segmentation by long term analysis of point trajectories, in: European Conference on Computer Vision, 2010, pp. 282–295.

[6] J. Lezama, K. Alahari, J. Sivic, I. Laptev, Track to the future: Spatio-temporal video segmentation with long-range motion cues, in: IEEE International Conference on Computer Vision and Pattern Recognition, 2011, pp. 3369–3376.

[7] H. Wang, A. Kläser, C. Schmid, C.-L. Liu, Action recognition by dense trajectories, IEEE International Conference on Computer Vision and Pattern Recognition (2011) 3169–3176.

[8] X. Cao, Z. Li, Q. Dai, Semi-automatic 2D-to-3D conversion using disparity propagation, IEEE Transactions on Broadcasting 57 (2) (2011) 491–499.

[9] N. Sundaram, T. Brox, K. Keutzer, Dense point trajectories by GPU-accelerated large displacement optical flow, European Conference on Computer Vision (2010) 438–451.

[10] H. Wang, A. Kläser, C. Schmid, C.-L. Liu, Dense trajectories and motion boundary descriptors for action recognition, International Journal of Computer Vision (2013) 1–20.

[11] M. Werlberger, W. Trobin, T. Pock, A. Wedel, D. Cremers, H. Bischof, Anisotropic Huber-L1 optical flow, British Machine Vision Conference.

[12] A. Salgado, J. Sánchez, Temporal constraints in large optical flow estimation, in: Computer Aided Systems Theory, 2007, pp. 709–716.

[13] P. Sand, S. Teller, Particle video: Long-range motion estimation using point trajectories, International Journal of Computer Vision 80 (1) (2008) 72–91.

[14] M. Irani, Multi-frame correspondence estimation using subspace constraints, International Journal of Computer Vision 48 (3) (2002) 173–194.

[15] R. Garg, A. Roussos, L. Agapito, A variational approach to video registration with subspace constraints, International Journal of Computer Vision 104 (3) (2013) 286–314.

[16] T. Crivelli, P.-H. Conze, P. Robert, P. Pérez, From optical flow to dense long term correspondences, in: IEEE International Conference on Image Processing, 2012.

[17] T. Crivelli, P.-H. Conze, P. Robert, M. Fradet, P. Pérez, Multi-step flow fusion: Towards accurate and dense correspondences in long video shots, in: British Machine Vision Conference, 2012.

[18] P.-H. Conze, T. Crivelli, P. Robert, L. Morin, Dense motion estimation between distant frames: Combinatorial multi-step integration and statistical selection, in: IEEE International Conference on Image Processing, 2013.

[19] P.-H. Conze, T. Crivelli, P. Robert, L. Morin, Dense long-term motion estimation via statistical multi-step flow, in: International Conference on Computer Vision Theory and Applications, 2014.

[20] J. Butcher, Numerical methods for ordinary differential equations, John Wiley & Sons, 2008.

[21] T. Crivelli, M. Fradet, P.-H. Conze, P. Robert, P. Pérez, Robust optical flow integration, IEEE Transactions on Image Processing 24 (1) (2015) 484–498.

[22] M. Rubinstein, C. Liu, W. T. Freeman, Towards longer long-range motion trajectories, in: British Machine Vision Conference, 2012.

[23] T. Brox, J. Malik, Large displacement optical flow: descriptor matching in variational motion estimation, IEEE Transactions on Pattern Analysis and Machine Intelligence 33 (3) (2011) 500–513.

[24] P. Robert, C. Thébault, V. Drazic, P.-H. Conze, Disparity-compensated view synthesis for s3d content correction, in: SPIE IS&T Electronic Imaging Stereoscopic Displays and Applications, 2012.

[25] R. Garg, A. Roussos, L. Agapito, Robust trajectory-space TV-L1 optical flow for non-rigid sequences, in: Energy Minimization Methods in Computer Vision and Pattern Recognition, 2011, pp. 300–314.

[26] A. Wedel, T. Pock, C. Zach, H. Bischof, D. Cremers, An improved algorithm for TV-L1 optical flow, in: Statistical and Geometrical Approaches to Visual Motion Analysis, 2009, pp. 23–45.

[27] D. Pizarro, A. Bartoli, Feature-based deformable surface detection with self-occlusion reasoning, International Journal of Computer Vision 97 (1) (2012) 54–70.

[28] R. Tron, R. Vidal, A benchmark for the comparison of 3D motion segmentation algorithms, in: IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1–8.