THE RECONSTRUCTION OF MISSING FRAMES IN HISTORICAL FILMS, A LAYERED APPROACH

Jan Fransens* Fabian Di Fiore† Frank Van Reeth‡

Limburgs Universitair Centrum
Expertisecentre for Digital Media
Universitaire Campus
B-3590 Diepenbeek, Belgium

Figure 1: From left to right: (a) and (b) The leading and trailing frames. (c) The layered representation. (d) The reconstruction of an intermediate frame. Notice the occlusions by the road sign and the human figure.

Abstract

Since the very beginning of cinematography, a historic wealth of information has been recorded and stored on film. Unfortunately, the very celluloid medium on which motion pictures are captured proves to be ultimately susceptible to severe degradation. These ageing affects should be reversed as much as possible. In addition, since manual restoration is a tedious and costly process, an automatic or semi automatic system is highly desirable. This paper addresses the most severe form of film degradation, in which complete sequences of frames are beyond repair by traditional restoration techniques, or are simply lost permanently. To this end, we propose a layered approach to the frame reconstruction problem, in which a reasonable amount of user interaction is allowed to completely resynthesize a given number of frames. We will demonstrate the results of our technique on a number of sequences from actual archive footage.

CR Categories: I.4.4 [Image Processing and Computer Vision]: Restoration—CINE - Synthesis

Keywords: film restoration, motion compensated frame interpolation, layered motion representation

*e-mail: jan.fransens@luc.ac.be
†e-mail: fabian.difiore@luc.ac.be
‡e-mail: frank.vanreeth@luc.ac.be

1 INTRODUCTION AND MOTIVATION

Classical film typically exhibits some forms of degradation, caused by the long time storage process. The celluloid medium on which movies haven been stored, causes archive material to be subject to several types of chemical and mechanical aberrations. Improper storage environments can contribute to such artifacts as dust and dirt, locally polluting the film material. Other degradations, such as mechanical abrasions, can affect an entire sequence of frames. Mold, which is also a common phenomenon in historic films, has an even more destructive impact on the film, and can mutilate entire sequences of frames.

All the material contained in the various film archives, constitutes a wealth of historic information, recording artistic and cultural developments spanning over a century, which is obviously essential to preserve. One solution is contained in the huge technological progress over the few last decades, enabling the transfer of archival material to a digital medium. In addition to digitisation, the affected material should also be restored to its original quality, preferably with an automatic approach.

As we will discuss in the next section, previous work on film reconstruction mostly focusses on repairing individual frame defects. In contrast, this paper addresses the most severe form of degradation, in which an entire sequence of frames is lost permanently, or degraded beyond local repair. Our contributions include a novel motion and feature adaptive mesh warping algorithm, which is tightly coupled with a user interactive layer segmentation. We will show how this hybrid approach results in a robust and general system for the reconstruction of missing film frames. We will demonstrate the results of our approach on a number of sequences from digitised archive film material.

2 RELATED WORK

There exists a large body of work on the restoration of degraded film material, most of which deals with the repair of individual frames, rather than the regeneration of lost frames. It is therefore not our intention, nor is it within the scope of this work, that a full treatise is given of these restoration techniques. We refer the interested reader to [Schallauer et al. 1999] and [Kokaram 1998], in which more general overviews on the topic of frame restoration are given.

Arguably the most common defect in historical film is the pollution of the frame content by dust and dirt which can be caused by an improper storage environment. Dust and dirt are manifested as the appearance of light or dark spots or *blotches*, which can be relatively easily removed with automated algorithms [Biemond et al. 1999] [Tenze et al. 2000]. Another frequently addressed problem is the automatic detection and removal of *line scratches* [Bretschneider et al. 2000] [Joyeux et al. 1999] [Besserer and Thire 2004], which are caused by the contact of the film with mechanical parts of the projector. These artifacts tend to be somewhat more problematic to handle, since rather than having a local effect, scratches can extend over a number of frames.

To the best of our knowledge, we are not aware of automatic frame regeneration techniques that specifically address the restoration of historical films. On the other hand, missing frame interpolation is a relatively well studied problem in the setting of video coding, and is used in video compression, where the high temporal redundancy of natural video sequences is exploited, as well as in other processing applications.

Frame rate conversion algorithms are concerned with retargeting a given sequence to a different frame rate. Applications include conversion to the 50Hz HDTV standard, or rendering a sequence in slow motion. Besides these applications, frame rate conversion is also commonly used in low bandwidth video coding, where a number of frames can be skipped [Han and Woods 1997] [Liu et al. 2000b] which are then interpolated during the decoding stage.

The simplest techniques employ simple schemes such as frame repetition and frame averaging [Tekalp 1995], which produce jerky motion and ghosting effects respectively, and are thus unsuitable for high end purposes.

To overcome these problems, more sophisticated frame interpolation approaches were developed, taking into account the motion trajectories between subsequent frames to arrive at a smoother transition. These motion trajectories are generated during the video encoding stage, and are derived from the motion estimated from individual pixel or entire block displacements. In either case, these motion vectors are likely to contain a certain degree of mismatched motion vectors. Outliers are filtered out to generate an approximation to the true motion vectors [Dane and Nguyen 2004] [Blume et al. 2002], after which intermediate frames are generated by proportionally warping the reference frame by the filtered motion vectors. Although there exist techniques that take occlusion effects into account [Liu et al. 2000a], the underlying assumption is that two subsequent frames are relatively close to each other with respect to their moving features. Also, the difficult case of ambiguous motion near object boundaries, is insufficiently handled.

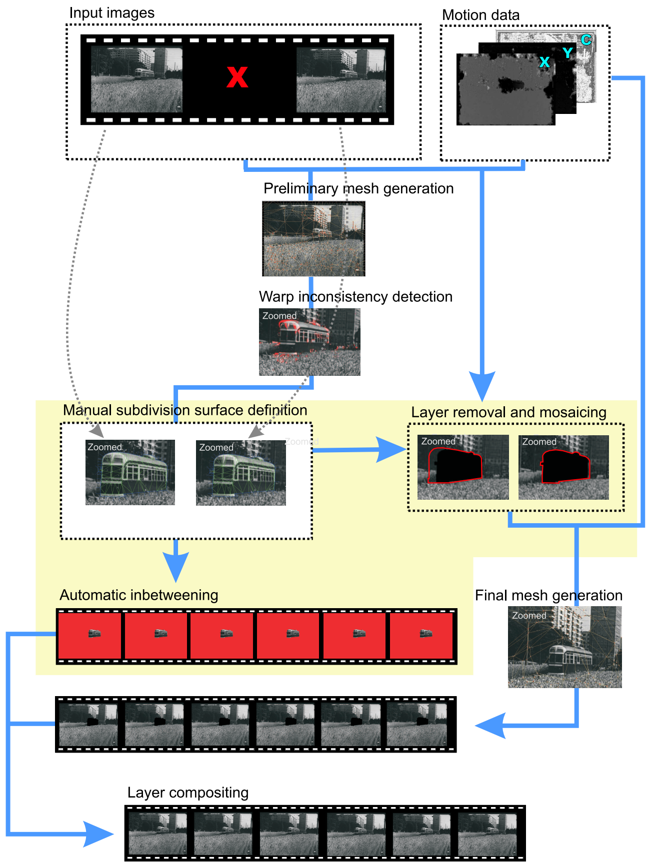

Figure 2: Schematic of the proposed frame interpolation framework.

A more unified way of dealing with occlusions and motion boundaries is described by the layer based motion description, in which the notion of independent motion layers within an image is introduced. This representation captures more of the salient features of real world motion, since an image is segmented into layers with depth ordering. Wang and Adelson [Wang and Adelson 1994] pioneered the layered motion representation by segmenting a dense optical flow field into regions of coherent motion, which defines the individual layers. With this layered description, a moving sequence can be described with a relatively small number of independently moving layers, each with their own velocity map. The major limitation of the approach lies in the simplicity of the motion segmentation, which assumes an affine motion model. Also, the segmentation into layers is imperfect as pixels with inaccurate motion vectors might be assigned to the wrong layer.

Subsequent work on layering address the segmentation accuracy with various methods such as graph cut segmentation [Xiao and Shah 2004] [Zhang et al. 2005], likelihood estimation [Jepson et al. 2002], or by imposing smoothness constraints [Nicolescu and Medioni 2003].

3 OVERVIEW

From the previous section, we conclude that the layered motion representation presents the most promising framework for the regeneration of an entire missing sequence. However, in the context of frame resynthesis, the correct ordering of the layers depends to great extent on the information that is available from the images that are used as a reference for the intermediate frame generation. Consider for example a sequence depicting an object travelling through a static scene, in which a scene object occludes the motion trajectory at some point of the sequence. Suppose that the missing frames are situated where normally the occlusions between the static and moving object are shown. In other words, the frames from which the synthesis is performed contain the moving object on either side of the occluder. In this situation, no single automatic algorithm will be able to correctly estimate the depth relationship between the moving object and the static scene.

In order to deal with complex layer interactions, we propose a semi automatic reconstruction framework, in which the notion of independent motion layers is adopted. Allowing a reasonable amount of user input, our system is more robust in terms of complex layer occlusions. In addition, our technique attains a greater flexibility in terms of the motion model employed for the resynthesis of the intermediate frames, in contrast with the simplified affine motion model often assumed in the various automatic motion segmentation algorithms.

Figure 2 shows an overview of the proposed system, which takes as input the two frames enclosing the missing frames together with a dense motion map, relating these reference frames. The core of the automatic algorithm consists of a mesh warping routine. A triangular mesh is superimposed on the reference frame, after which the nodal points are displaced according to their associated motion vectors which are derived from a dense motion map, thereby deforming the underlying content to generate intermediate frames. The topology of the mesh is a crucial factor in the quality of the reconstructed frames, and the nodal points should preferably be concentrated on strong image features. The automatic mesh design is discussed in more detail in the next section.

User intervention involves the layer specification, in terms of both content and relative ordering. A specialised subdivision surface is defined over each object of interest in both reference frames, so that the vertices reside on corresponding object features. These pairs of related subdivision surfaces each define a motion layer, for which the user can indicate the relative ordering. Once the motion of independently moving objects is segmented using this procedure, in-betweens are generated for each separate layer, guided by the two interrelated meshes. During this process, the user can easily correct the nodal positions of each intermediate mesh, to ensure a physically plausible deformation.

The final stage of the frame reconstruction involves the compositing of each of the layered in-betweens, in accordance with the imposed order, with the automatically generated background layer.

4 BACKGROUND MESH DESIGN

The background layer is basically treated as an elastic deformable sheet, which is realised by means of a two dimensional mesh. Intermediate frames are generated by displacing the vertices with a fraction of the associated motion vectors. More formally, if frames J and $J+N$ are the leading and trailing frames respectively, the intermediate frame $J+n$ is generated by displacing the nodal points (x_i, y_i) with associated motion (u_i, v_i) according to the following relationship:

$$\begin{aligned} x_i' &= x_i + \frac{n}{N} u_i \\ y_i' &= y_i + \frac{n}{N} v_i \end{aligned}$$

The mesh topology is essential for the quality of the generated intermediate frames, and should be adaptive to both the image features and the underlying motion field. To this end, we impose three criteria for the design of the mesh:

- The nodal points should preferably be placed on strong image features by using a measure such as the image gradient.
- The mesh should be adaptive to the underlying motion, that is, in regions with a large variation in motion direction and size, more mesh vertices should be placed with respect to those regions exhibiting a more uniform motion distribution.
- The motion vectors associated with the nodal points should have a high accuracy, to ensure the correct warped position.

Figure 3: Motion adaptive warping mesh, superimposed on a reference frame from a panning sequence. By weighing the candidate vertices with the image gradient, nodal points are placed on strong image features.

These three constraints are addressed with an adaptive quadtree approach in combination with a motion confidence map. This confidence map is introduced, because the dense motion estimation is very likely to be affected by the noise contained in historical film material, even after manual or automatic cleanup. The motion confidence C gives per motion vector an accuracy likelihood, and is constructed by the normalised cross correlation of every pixel in one reference frame with its motion compensated counterpart in the other frame.

The quadtree is constructed using the distribution of the underlying motion field as a splitting criterion. At every level of the recursion, σ_v and $\sigma_{|v|}$ are determined for the quadrant under consideration, denoting the standard deviations of the motion vector direction and magnitude respectively. If either of the standard deviations do not satisfy the threshold τ, the quadrant is uniformly split. In the other case, thereby ensuring a uniform motion field is underlying the quadrant, a point within this quadrant is chosen as a nodal point. For each pixel (x_i, y_i) in the quadrant, a weight w_i is calculated by multiplying the motion confidence of the pixel's associated motion vector with the magnitude of the image gradient G:

$$w_i = b(x_i, y_i) \cdot C(x_i, y_i) \cdot \|G(x_i, y_i)\|_2$$

where b is a binary valued function, returning 0 if (x_i, y_i) lies within a distance r of a previously generated nodal point. The point (x_i, y_i) that maximises the weight w_i is chosen as a mesh vertex.

Preceding the actual mesh generation, a number of uniformly spaced nodal points along the image borders are added to the points

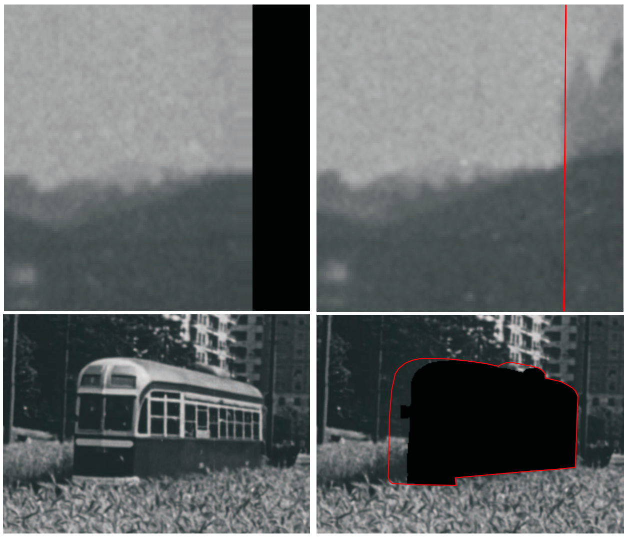

Figure 4: Reference frame mosaicing for image completion. Aligning the two leading and trailing frames into a common reference frame, missing data due to inward warping (top) or layer extraction (bottom) can be partially filled in.

generated by the quadtree approach, to ensure complete image coverage of the mesh. The final mesh is constructed using a Delaunay triangulation.

The inaccuracy of motion estimation along image borders is a typical problem with dense motion estimation. Moreover, digitised historic films often show a black border surrounding the actual film frame as a result of the scanning process (for instance, see figure 3). Since we have appended nodal points along these image borders, we must derive accurate motion vectors instead of relying on the dense motion map. We do this by averaging the motion vectors of connected nodal points in the Delaunay graph. This propagates the accurate motion of the vertices generated with the quadtree approach to the border vertices, ensuring no stretching or folding artifacts occur during the warping to intermediate frames.

Since we are warping from one reference frame to interpolate the motion to the other frame, the motion of a border point may be such that the warped frame does not completely cover the entire image. By mosaicing [Shum and Szeliski 1997] the two reference frames, we obtain a complete image of the background. This mosaiced frame is then used for generating intermediate frames (figure 4).

5 LAYERED 2D MESH BASED MORPHING

The morphing or metamorphosis of images or image parts is often used to generate special effects in films and animation. Within the setting of layered in-betweening, we use an adaptation of the basic notion of warping.

Previous work in mesh based morphing, as described in [Wolberg 1990], assumes a uniform mesh of points in the source image, which define a grid of control points. The correspondence with a target image is defined, by manually adjusting these control points to match the features between frames. Subsequent work addresses the tedious manual and error-prone feature relating, for instance by automating the process with active nets [Nishita et al. 1993],snakes [Lee et al. 1996], by using pairs of directed line segments [Lee et al. 1998], or by specifying outline curves [Tal and Elber 1999][Johan et al. 2000].

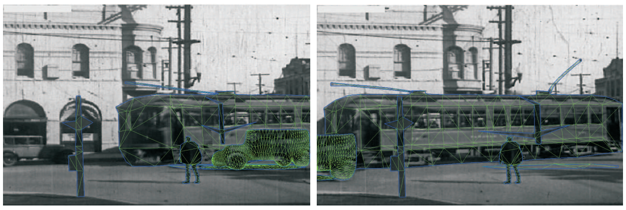

Figure 5: Layered mesh specification. For each motion layer, a pair of corresponding meshes is defined in the reference frames, thereby establishing a layered motion trajectory.

The general problem of interpolating between two dimensional shapes is still an active domain of research. For our purposes, we adapt the approach of [Van den Bergh et al. 2002], in which the morphing between general objects with concavities, holes or protruding limbs is specifically addressed. Moreover, local control can be maintained by the use of subdivision techniques.

For each of the corresponding objects of interest, a motion layer is defined by segmenting the objects with an enclosing polygon. This polygon is then triangulated and refined with locally interpolating subdivision techniques. Note that this mesh is created for only one reference frame. The motion layer correspondence is established by placing the same mesh on the other reference frame, and manually adjusting the vertices to match the image features. Figure 5 shows an example of the motion layer mesh pairs.

Using the layer correspondences, intermediate frame layers are automatically generated by piecewise affine warping, as in the previous section. Intermediate frames can be further refined by the manual adjustment of vertices, thereby allowing more general object transformations.

6 COMPOSITING

The final step of the frame reconstruction consist of combining the various in-betweens of the layers into a composite sequence. After the background layer warps are generated with the quadtree mesh, the interpolations of the user defined layers are composited with these frames, in the order specified during the layer modelling step . Note that the background warping uses the mosaiced reference frames (see figure 4), for which the confidence map is set to zero for pixels which are occluded by a user defined layer. This way, the quadtree mesh design algorithm does not generate any nodal points on pixels belonging to these layers. The mosaicing step ensures that during the final compositing no borders of the cut out regions of the background layer are shown.

In order to smoothly blend the various layers onto a final frame, a simple alpha map is created for each segmented layer, by convolving the binary mask of that layer with a Gaussian kernel.

7 RESULTS

In figure 6 we present the results of our approach on a number of sequences from digitised archive footage. The examples are ordered in order of complexity and amount of user interaction. The respective animated versions of the results can be downloaded from `http://research.edm.luc.ac.be/~fdifiore/research/GraphiCon2005/`.

The first example depicts the results for a simple panning sequence. In this example, 6 frames were fully automatically resynthesized, by warping the background layer. The resynthesis of the

Figure 6: The results of our algorithm. The first and fourth row show the leading and trailing reference frames respectively. For each sequence, two intermediate frames out of 6(two leftmost columns) and 41(right column) missing frames are shown.

second example, for which the reference frames are also 6 frames apart, was generated by defining an extra layer containing the the tramcar, moving in the opposite direction of the camera pan. For the last sequence, a total of 41 frames were reconstructed from the two reference frames and the definition of 3 layers, composited on the background. As can be seen from the figure, the recovered intermediate frames show the correct occlusion ordering.

8 CONCLUSIONS AND FUTURE WORK

In this work, we have presented a layered approach for the complete resynthesis of entire sequences of historic film material. To this end, a hybrid system is proposed, with a tightly coupled interaction between fully automated reconstruction and manual intervention. A scene is decomposed into a number of motion layers, by manually establishing correspondences between objects of interest in the leading and trailing frame. Using this information, the motion layers are independently morphed and composited onto the automatically generated background layer. The background is piecewise affine warped, guided by a mesh structure that adapts itself to the underlying motion and the image features, in addition to an estimation of the relative motion accuracy. We have demonstrated the results of our approach on a number of actual archive films.

Future work includes incorporating the analysis of motion trajectories over a number of preceding and succeeding frames with respect to the lost material to generate more general warping paths. Also, the polygonal segmentation step may be improved on by employing more powerful segmentation techniques, such as graphcut, to refine the segmented boundaries.

ACKNOWLEDGEMENTS

We gratefully express our gratitude to the European Fund for Regional Development and the Flemish Government, which are kindly funding part of the research reported in this paper.

Part of the work is also funded by the European research project IST-2001-37117 'RACINE-S'. In this context, we would like to express our thanks to the various people in the project providing ideas and for helping with the implementation.

We are also deeply indebted to Johan Claes for his valuable paper reviews.

REFERENCES

BESSERER, B., AND THIRE, C. 2004. Detection and tracking scheme for line scratch removal in an image sequence. In *Proceedings of ECCV 2004*, 264–275.

BIEMOND, J., VAN ROOSMALEN, P., AND LAGENDIJK, R. 1999. Improved blotch detection by postprocessing. In *Proceedings of ICASSP 1999*.

BLUME, G., HERCZEG, G., ERDLER, O., AND NOLL, T. 2002. Object based refinement of motion vector fileds applying probabilistic homogenization rules. *IEEE Transactions on Consumer Electronics 48*, 3 (August), 694–701.

BRETSCHNEIDER, T., KAO, O., AND BONES, P. 2000. Removal of vertical scratches in digitised historical film sequences using wavelet decomposition. In *Proceedings of the Image and Vision Computing New Zealand*, 37–42.

DANE, G., AND NGUYEN, T. 2004. Motion vector processing for frame rate up conversion. In *Proceedings of IEEE 2004 International Conference on Acoustics, Speech and Signal Processing*.

HAN, S., AND WOODS, J. 1997. Frame-rate up-conversion using transmitted motion and segmentation fields for very low bit-rate video coding. In *Proceedings of ICIP 1997*, 747–750.

JEPSON, A., FLEET, D., AND BLACK, M. 2002. A layered motion representation with occlusion and compact spatial support. In *Proceedings of ECCV 2002*, 692–706.

JOHAN, H., KOISO, Y., AND NISHITA, T. 2000. Morphing using curves and shape interpolation techniques. In *Proceedings of Pacific Graphics 2000, 8th Pacific Conference on Computer Graphics and Applications*, 348–358.

JOYEUX, L., BUISSON, O., BESSERER, B., AND BOUKIR, S. 1999. Detection and removal of line scratches in motion picture films. In *Proceedings of CVPR 1999*, 548–553.

KOKARAM, A. 1998. *Motion Picture Restoration: Digital Algorithms for Artefact Suppression in Degraded Motion Picture Film and Video*. Springer Verlag, ISBN: 3-540-76040-7.

LEE, S., WOLBERG, G., CHWA, K.-Y., AND SHIN, S. Y. 1996. Image metamorphosis with scattered feature constraints. *IEEE Transactions on Visualization and Computer Graphics 2*, 14, 337–354.

LEE, S., CHWA, K.-Y., HAHN, J., AND SHIN, S. Y. 1998. Fast feature-based volume metamorphosis and operator design. In *Computer Graphics Forum*, vol. 17(3), 15–22.

LIU, S., KIM, J., AND KUO, C. 2000. MCI-embedded motion compensated prediction for quality enhancement of frame interpolaton. In *SPIE Proceedings of International Symposium on Voice, Video, and Data Communications, Multimedia Systems and Applications III*.

LIU, S., KIM, J., AND KUO, C. 2000. Non-linear motion-compensated interpolation for low bit rate video. In *SPIE Proceedings of International Symposium on Optical Science, Engineering, and Instrumentation 2000, Applications Digital Image Processing XXIII*.

NICOLESCU, M., AND MEDIONI, G. 2003. Layered 4d representation and voting for grouping from motion. *PAMI 25*, 4 (April), 492–501.

NISHITA, T., FUJII, T., AND NAKAMAE, E. 1993. Metamorphosis using Bézier clipping. In *Proceedings of 1st Pacific Conference on Computer Graphics and Applications*, 162–173.

SCHALLAUER, P., PINZ, A., AND HAAS, W. 1999. Automatic restoration algorithms for 35mm film. *Videre 1*, 3, 60–85.

SHUM, H., AND SZELISKI, R. 1997. Panoramic image mosaics. Tech. Rep. MSR-TR-97-23, Microsoft Research.

TAL, A., AND ELBER, G. 1999. Image morphing with feature preserving texture. In *Proceedings of Eurographics (EG1999)*, vol. 18(3), 339–348.

TEKALP, A. 1995. *Digital Video Processing*. Prentice-Hall.

TENZE, L., RAMPONI, G., AND CARRATO, S. 2000. Blotches correction and contrast enhancement for old film pictures. In *Proceedings of International Conference on Image Processing 2000*, 660–663.

VAN DEN BERGH, J., DI FIORE, F., CLAES, J., AND VAN REETH, F. 2002. Interactively morphing irregularly shaped images employing subdivision techniques. In *Proceedings of 1st Ibero-American Symposium in Computer Graphics (SIACG2001)*, 315–321.

WANG, J. Y., AND ADELSON, E. H. 1994. Representing moving images with layers. *IEEE Transactions on Image Processing Special Issue: Image Sequence Compression 3*, 5 (September), 625–638.

WOLBERG, G. 1990. *Digital Image Warping*. IEEE Computer Society Press, ISBN: 0-8186-8944-7.

XIAO, J., AND SHAH, M. 2004. Motion layer extraction in the presence of occlusion using graph cut. In *Proceedings of CVPR 2004*, 972–979.

ZHANG, Y., XIAO, J., AND SHAH, M. 2005. Motion layer based object removal in videos. In *Proceedings of the IEEE Workshop on Applications of Computer Vision*. to appear.

About the Authors

Jan Fransens is a Ph. D. student of computer science at Limburgs Universitair Centrum, Expertise Centre for Digital Media (EDM).

His contact email is jan.fransens@luc.ac.be.

Fabian Di Fiore is a postdoctoral research associate of computer science at Limburgs Universitair Centrum, Expertise Centre for Digital Media (EDM). His contact email is fabian.difiore@luc.ac.be.

Frank Van Reeth is a professor of computer science at Limburgs Universitair Centrum and deputy managing director of the Expertise Centre for Digital Media (EDM). His contact email is frank.vanreeth@luc.ac.be.