Video Completion by Motion Field Transfer

Takaaki Shiratori*  Yasuyuki Matsushita**  Sing Bing Kang†  Xiaoou Tang**
* The University of Tokyo ** Microsoft Research Asia † Microsoft Research

→ Why Use Motion Field?
- Color-based method: Requires similar color & motion
+ Motion-based method: Requires only similar motion

More chance to fill-in a hole!

Motion can be copied from video portions with different appearance.

→ Motion Dissimilarity

\[ m = (ut, vt, t)^T \] : motion vector in 3D spatio-temporal domain 
(t : frame interval)

Motion dissimilarity

\[ d_m(m_0, m_1) = 1 - \frac{m_0 \cdot m_1}{||m_0|| \cdot ||m_1||} = 1 - \cos \theta \]

\( \theta \) : Angular distance between \( m_0 \) and \( m_1 \) in homogeneous coordinates

→ Motion Field Transfer

An optimal source patch \( P_t \) minimizes

\[ d(P_t(x_i), P_t(x_i)) = \frac{1}{|D|} \sum_{p \in D} d_m(m(x_i + p), m(x_i + p)) \]

Find optimal source patches in coarse-to-fine manner.

→ Color Propagation

Color of \( \text{color } c_i \) is estimated from \( \text{color } c \)

Weighting factor for each patch

\[ w = \frac{s}{d} \]

\( s \) : overlapped area
\( d \) : patch distance

\[ c(\bullet) = \sum_{w \in W} wc(\bullet) \]

For all hole pixels,

\[ C = [W \mid W_b] \begin{bmatrix} C \hline C_b \end{bmatrix} \]

\( C_b \): color of hole pixels
\( C_b \): color of boundary pixels (already defined)

\[ C = (I - W)^{-1} W_b C_b \]

sparse & structurally symmetric matrix
Experimental Results

Hole-filling

| Ground truth | Spatio-temporal hole (magenta) and computed optical flow (green) | Result of motion field transfer | Result of video completion |

Object removal

| Original video | Foreground person is removed. | Completed video |

Frame interpolation

| recovered frame |

Frame rate recovery using the motion field prior

| Prior video | Output video |

Discussion

Color-based method vs. Motion-based method

| (a) Ground truth | (b) Result of our method | (c) Intensity difference between (a) and (b) | (d) Result of color-based method | (e) Intensity difference between (a) and (d) |

Blur effects

| Ground truth |

Temporal size 2 5 7 10