Modifying Non-Local Variations Across Multiple Views

Tal Tlusty
Technion, Israel
taltlusty@gmail.com

Tomer Michaeli
Technion, Israel
tomer.m@ee.technion.ac.il

Tali Dekel
Google Research
tdekel@google.com

Lihi Zelnik-Manor
Technion, Israel
lihi@technion.ac.il

1. APPENDIX A: Derivation of the transformation update step under the transformation consistency loss

In this Appendix, we start by briefly reviewing the formulation underlying the optical flow algorithm of Liu [3]. We then extend this formulation to our multi-view setting by adding the correspondence term

$$E_{\text{corr}}(\mathcal{T}_1, \mathcal{T}_2) = \iint \left(|u_1(x,y) - u_2(x,y)|^2 + |v_1(x,y) - v_2(x,y)|^2\right) dxdy, \tag{1}$$

which forces small differences between the transformations $\mathcal{T}_1$ and $\mathcal{T}_2$.

1.1. Single image optical-flow

Let $p = (x, y)$ denote the image coordinates and $F_1(p) = (u_1(p), v_1(p))$ the flow field between image J_1 and image I_1. The optical flow objective function addressed in [3] is given by

$$E_{\text{OF}}(u, v) = \int \psi(|I_1(p + F_1(p)) - J_1(p)|^2)dp + \alpha \int \phi(|\nabla u_1(p)|^2 + |\nabla v_1(p)|^2)dp. \tag{2}$$

In our case, we are interested in the robust penalty functions $\psi(z) = \sqrt{z^2 + \epsilon^2}$ and $\phi(z) = \sqrt{z^2 + \epsilon^2}$.

Under the incremental flow framework, assuming some initial estimate F_1 of the flow field is available, the goal is to find the best increment $dF_1 = (du, dv)$. The objective function in (2) can thus be written as

$$\begin{aligned} E_{\text{OF}}(du, dv) = &\int \psi(|I_1(p + F_1(p) + dF_1(p)) - J_1(p)|^2)dp \\ &+ \alpha \int \phi(|\nabla(u_1(p) + du_1(p))|^2 + |\nabla(v_1(p) + dv_1(p))|^2)dp \end{aligned} \tag{3}$$

Let $I_{z,1}(p) = I_1(p + F_1(p)) - J_1(p)$, $I_{x,1}(p) = \frac{\partial}{\partial x} I_1(p + F_1(p))$, and $I_{y,1}(p) = \frac{\partial}{\partial y} I_1(p + F_1(p))$. Then the term $I_1(p + F_1(p) + dF_1(p)) - J_1(p)$ can be linearized by a first order Taylor expansion as follows

$$I_1(p + F_1(p) + dF_1(p)) - J_1(p) \approx I_{z,1}(p) + I_{x,1}(p)du_1(p) + I_{y,1}(p)dv_1(p). \tag{4}$$

Vectorizing $u_1, v_1, du_1, dv_1, I_{x,1}, I_{y,1}, I_z$ into $U_1, V_1, dU_1, dV_1, \tilde{I}_{x,1}, \tilde{I}_{y,1}, \tilde{I}_{z,1}$, the energy function in (3) can be discretized as

$$E_{\text{OF}}(dU_1, dV_1) = \sum_p \psi(f_{p,1}) + \alpha \sum_p \phi(g_{p,1}), \tag{5}$$

where

$$f_{p,1} = (\delta_p^T(\tilde{I}_{z,1} + \mathbf{I}_{x,1}dU_1 + \mathbf{I}_{y,1}dV_1))^2 \tag{6}$$

$$g_{p,1} = (\delta_p^T(\mathbf{D}_x(U_1 + dU_1)))^2 + (\delta_p^T(\mathbf{D}_y(U_1 + dU_1)))^2 + (\delta_p^T(\mathbf{D}_x(V_1 + dV_1)))^2 + (\delta_p^T(\mathbf{D}_y(V_1 + dV_1)))^2 \tag{7}$$

Here, δ_p is a vector with 1 at the pth location and 0 everywhere else, $\mathbf{D}_x, \mathbf{D}_y$ are matrices that correspond to convolutions with the horizontal and vertical derivative filters $[-1\,1]$, $[-1\,1]^T$, respectively, and $\mathbf{I}_{x,1} = \text{diag}(\tilde{I}_{x,1})$, $\mathbf{I}_{y,1} = \text{diag}(\tilde{I}_{y,1})$.

1.2. Multi-view optical-flow

In our multiview setting, we are seeking two flow fields. Our approach is to alternate between fixing the second flow field and solving for the first, and vice versa. Without loss of generality, let us regard the second flow field U_2, V_2 as constant as solve for dU_1, dV_1.

Adding the discretized correspondence term of (1) to (5), the multi-view optical flow loss for updating dU_1, dV_1 can be written as

$$E_{\text{MV-OF}}(dU_1, dV_1) = \sum_p \psi(f_{p,1}) + \alpha \sum_p \phi(g_{p,1}) + \alpha_c^* \sum_p t_{p,1}, \tag{8}$$

where

$$t_{p,1} = (\delta_p^T(U_1 + dU_1 - U_2))^2 + (\delta_p^T(V_1 + dV_1 - V_2))^2 \tag{9}$$

and $\alpha = \frac{\alpha_r}{\lambda}, \alpha_c^* = \frac{\alpha_c}{\lambda}$.

The key idea behind the Iterative Re-weighted Least Squares (IRLS) method [1] is to linearize the functions ψ and ϕ and regard their gradients as constant in each iteration. In our case,

$$\nabla_{U_1} E_{\text{MV-OF}}(dU_1, dV_1) = \sum_p \psi^{'}(f_{p,1}) \nabla_{U_1} f_{p,1} + \alpha \sum_p \phi^{'}(g_{p,1}) \nabla_{U_1} g_{p,1} + \alpha_c^* \sum_p \nabla_{U_1} t_{p,1}, \tag{10}$$

where

$$\nabla_{U_1} f_{p,1} = 2(\mathbf{I}_{x,1}\delta_p\delta_p^T\mathbf{I}_{x,1} dU_1 + \mathbf{I}_{x,1}\delta_p\delta_p^T(I_{z,1} + \mathbf{I}_{y,1} dV_1)), \tag{11}$$

$$\nabla_{U_1} \phi_{p,1} = 2((\mathbf{D}_x^T\delta_p\delta_p^T\mathbf{D}_x + \mathbf{D}_y^T\delta_p\delta_p^T\mathbf{D}_y)(dU_1 + U_1)), \tag{12}$$

$$\nabla_{U_1} t_{p,1} = 2(\delta_p\delta_p^T)(dU_1 + U_1 - U_2). \tag{13}$$

A similar expression can be derived for $\nabla_{V_1} E_{\text{MV-OF}}(dU_1, dV_1)$. Equating $\nabla_{U_1} E_{\text{MV-OF}}$ and $\nabla_{V_1} E_{\text{MV-OF}}$ to zero and reorganizing yields

$$(\Psi_1^{'}\mathbf{I}_{x,1}^2 + \alpha\mathbf{L}_1 + \alpha_c^*\mathbf{Id})dU_1 + \Psi_1^{'}\mathbf{I}_{y,1}\mathbf{I}_{x,1} dV_1 = -\Psi_1^{'}\mathbf{I}_{x,1} I_{z,1} - (\alpha\mathbf{L}_1 + \alpha_c^*\mathbf{Id})U_1 + \alpha_c^* U_2 \tag{14}$$

$$\Psi_1^{'}\mathbf{I}_{x,1}\mathbf{I}_{y,1} dU_1 + (\Psi_1^{'}\mathbf{I}_{y,1} + \alpha\mathbf{L}_1 + \alpha_c^*\mathbf{Id})dV_1 = -\Psi_1^{'}\mathbf{I}_{y,1} I_{z,1} - (\alpha\mathbf{L}_1 + \alpha_c^*\mathbf{Id})V_1 + \alpha_c^* V_2 \tag{15}$$

where $\mathbf{Id}$ is the identity matrix and $\mathbf{L}_1 = \mathbf{D}_x^T\Phi_1^{'}\mathbf{D}_x + \mathbf{D}_y^T\Phi_1^{'}\mathbf{D}_y$, with $\Phi_1^{'}$ and $\Psi_1^{'}$ denoting diagonal matrices with $\{\phi'(g_{p,1})\}$ and $\{\psi'(f_{p,1})\}$ on their diagonals. Writing this system of equations in matrix form, we get

$$\begin{pmatrix} \Psi_1^{'}\mathbf{I}_{x,1}^2 + \alpha\mathbf{L}_1 + \alpha_c^* & \Psi_1^{'}\mathbf{I}_{y,1}\mathbf{I}_{x,1} \\ \Psi_1^{'}\mathbf{I}_{y,1}\mathbf{I}_{x,1} & \Psi_1^{'}\mathbf{I}_{y,1}^2 + \alpha\mathbf{L}_1 + \alpha_c^* \end{pmatrix} \begin{pmatrix} dU_1 \\ dV_1 \end{pmatrix} = -\begin{pmatrix} \Psi_1^{'}\mathbf{I}_{x,1} I_{z,1} + \alpha\mathbf{L}_1 U_1 + \alpha_c^*(U_1 - U_2) \\ \Psi_1^{'}\mathbf{I}_{y,L} I_{z,1} + \alpha\mathbf{L}_1 V_1 + \alpha_c^*(V_1 - V_2) \end{pmatrix}$$

The idea in IRLS is to regard Φ', Ψ' as fixed (computed from the flow in the previous iteration), and obtain dU_1, dV_1 as the solution to this linear set of equations. Then Φ', Ψ' are updated based on the new dU_1, dU_2, leading to a new set of equations, etc.

Once U_1, V_1 are determined, we keep them fixed and update U_2, V_2 in a similar manner.

2. Appendix B : Appearance consistency

Recall that the overall multiview NLV objective is defined as

$$E_{\text{MV-NLV}}(\mathcal{T}_1, J_1, \text{DB}_1, \mathcal{T}_2, J_2, \text{DB}_2) = E_{\text{NLV}}(\mathcal{T}_1, J_1, \text{DB}_1) + E_{\text{NLV}}(\mathcal{T}_2, J_2, \text{DB}_2) + \alpha_c E_{\text{corr}}(\mathcal{T}_1, \mathcal{T}_2, J_1, J_2). \tag{16}$$

The correspondence term to maintain appearance consistency across views is

$$\begin{aligned}&E_{\text{corr}}(\mathcal{T}_1, J_1, \mathcal{T}_2, J_2) = \\ &\tfrac{\lambda}{\alpha_c} \iint \psi\left(\|\mathcal{T}_1^{-1}\{J_1\}(x,y) - \mathcal{T}_2^{-1}\{J_2\}(x,y)\|^2\right) dxdy + \iint \psi\left(\|\nabla w_x(x,y)\|^2 + \|\nabla w_y(x,y)\|^2\right) dxdy,\end{aligned} \tag{17}$$

where $w_x(x,y) = u_1(x,y) - u_2(x,y)$ and $w_y(x,y) = v_1(x,y) - v_2(x,y)$.

Since $\mathcal{T}_1\{I_1\}(x,y) = I_1(x+u_1(x,y), y+v_1(x,y))$ and the flow field u_1, v_1 is smooth, we can approximate $\mathcal{T}_1^{-1}\{J_1\}(x,y) \approx J_1(x - u_1(x,y), y - v_1(x,y))$. Similarly, $\mathcal{T}_2^{-1}\{J_2\}(x,y) \approx J_2(x - u_2(x,y), y - v_2(x,y))$.

2.1. Image update

As we show later on in Appendix C, setting the gradient of the objective w.r.t. J_1 to zero, leads to

$$J_1(x,y) = \gamma_1(x,y) Z_1(x,y) + \delta_1(x,y) I_1^c(x,y) + (1 - \gamma_1(x,y) - \delta_1(x,y)) J_{2\to1}(x,y), \tag{18}$$

where $J_{2\to1}$ is the image J_2 warped to the coordinate system of J_1, and

$$\begin{aligned}\gamma_1(x,y) &= \frac{W^1_{\text{data}}(x,y)}{W^1_{\text{data}}(x,y) + \frac{h^2}{M^2} + \frac{h^2}{M^2}\frac{W^1_{\text{data}}(x,y)}{M^1_{\text{data}}(x,y)} C(x,y)}, \\ \delta_1(x,y) &= \frac{\frac{h^2}{M^2}}{W^1_{\text{data}}(x,y) + \frac{h^2}{M^2} + \frac{h^2}{M^2}\frac{W^1_{\text{data}}(x,y)}{M^1_{\text{data}}(x,y)} C(x,y)},\end{aligned} \tag{19}$$

with $M^1_{\text{data}}(x,y) = \frac{1}{\lambda}\psi(\|J_1(x,y) - J_{2\to1}(x,y)\|^2)$.

Here, Z_1 is an image obtained by replacing each patch in J_1 by a weighted combination of its K Nearest Neighbor (NN) patches from the database DB_1, $W^1_{\text{data}}(x,y) = \frac{1}{\lambda}\psi(\|J_1(x,y) - I_1^c(x,y)\|^2)$, and C is an occlusion mask indicating for which pixels in I_2 we could find a good match in I_1 with high confidence. Since Z_1, γ_1 and δ_1 are nonlinear functions of the unknown J_1, we iterate between computing J_1 according to (18), and updating γ_1 and δ_1 according to (19). This whole process is done while regarding $J_{2\to1}$ as constant. The update of J_2 is done similarly, by regarding $J_{1\to2}$ as constant.

2.2. Transformation update

For this formulation, the term E_{corr} contains two parts which depend on $\mathcal{T}$, as described below:

$$\begin{aligned}&E(u_1, v_1, u_2, v_2) = \\ &\lambda \iint \psi(\|\mathcal{T}_1^{-1} J_1(x,y) - I_1(x,y)\|^2) + \psi(\|\mathcal{T}_2^{-1} J_2(x,y) - I_2(x,y)\|^2) + \psi(\|\mathcal{T}_1^{-1} J_1(x,y) - \mathcal{T}_2^{-1} J_2(x,y)\|^2) dxdy \\ &\qquad + \alpha_r \iint \psi(\|\nabla u_1(x,y)\|^2 + \|\nabla v_1(x,y)\|^2) dxdy + \alpha_c \iint \psi(\|\nabla w_x(x,y)\|^2 + \|\nabla w_y(x,y)\|^2) dxdy.\end{aligned} \tag{20}$$

Here (w_x, w_y) is the correction field that is added to the second view. We minimize this objective w.r.t. to $\mathcal{T}_1$ and $\mathcal{T}_2$ by extending the IRLS method [3]. See Appendix D for detailed description.

3. APPENDIX C : Derivation of the image update step under the appearance consistency loss

We present the mathematical procedure for the first view, a similar procedure is applied to the second view. Minimizing the objective function

$$\begin{aligned} &E_{\text{MV-NLV}}(\mathcal{T}_1, J_1, \text{DB}_1, \mathcal{T}_2, J_2, \text{DB}_2) = \\ &\qquad E_{\text{NLV}}(\mathcal{T}_1, J_1, \text{DB}_1) + E_{\text{NLV}}(\mathcal{T}_2, J_2, \text{DB}_2) + \alpha_c E_{\text{corr}}(\mathcal{T}_1, \mathcal{T}_2, J_1, J_2), \end{aligned} \tag{21}$$

with respect to J_1 requires setting the gradient of $E_{\text{rec}}(J_1, \text{DB}_1) + \lambda E_{\text{data}}(\mathcal{T}_1\{I_1\}, J_1) + \alpha_c E_{\text{corr}}(\mathcal{T}_1^{-1}\{J_1\}, \mathcal{T}_2^{-1}\{J_2\})$ with respect to J_1 to zero. In the following, I_1^c and J_1 denote the column-vector representations of $\mathcal{T}_1\{I_1\}$ and $J_1(x,y)$, respectively. We can express the recurrence term and the data term as

$$E_{\text{rec}}(J_1, \text{DB}_1) = -\sum_j \log\left(\sum_i \exp\left(-\tfrac{1}{2h^2}\|Q_j J - q_{1,i}\|^2\right)\right), \tag{22}$$

$$E_{\text{data}}(\mathcal{T}_1\{I_1\}, J_1) = \sum_k \psi((\delta_k^T(J_1 - I_1^c))^2), \tag{23}$$

where $q_{1,i}$ is the ith patch in DB_1, Q_j is the matrix that extracts jth $M \times M$ patch from the image, and δ_k is as in (7). The correspondence term that depends on J_1 is:

$$E_{\text{corr}}(\mathcal{T}_1, J_1, \mathcal{T}_2, J_2) = \lambda \iint \psi\left(\|\mathcal{T}_1^{-1}\{J_1\}(x,y) - \mathcal{T}_2^{-1}\{J_2\}(x,y)\|^2\right) dxdy \tag{24}$$

Assuming a smooth flow field (u, v) yields:

$$\begin{aligned} \mathcal{T}_1^{-1}\{J_1\}(x,y) - &\mathcal{T}_2^{-1}\{J_2\}(x,y) \\ &\approx J_1(x - u(x,y), y - v(x,y)) - J_2(x - u(x,y) + w_x(x,y), y - v(x,y) + w_y(x,y)) \\ &\approx J_1(x,y) - J_2(x + w_x(x,y), y + w_y(x,y)). \end{aligned} \tag{25}$$

To simplify the exposition we define $J_{2\to1}(x,y) = J_2(x + w_x(x,y), y + w_y(x,y))$. Substituting (25) into (24) and discretizing, we get

$$E_{\text{corr}}(\mathcal{T}_1, J_1, \mathcal{T}_2, J_2) = \lambda \sum_k \psi\left(C_k(\delta_k^T(J_{2\to1} - J_1))^2\right), \tag{26}$$

where $C_k = 1$ if the kth pixel has been warped correctly with high precision and $C_k = 0$ otherwize. We compute the gradient of the above components with respect to J_1 and set the total sum to zero.

3.1. Derivations

As shown in [2],

$$\nabla_{J_1} E_{\text{rec}} = \tfrac{1}{h^2} \sum_j Q_j^T Q_j J_1 - \tfrac{M^2}{h^2} Z_1, \tag{27}$$

where, $Z_1 = \frac{1}{M^2}\sum_j Q_j^T z_{1,j}$, $z_{1,j} = \sum_k w_{kj} q_{1,k}$, and $w_{kj} = \frac{\exp(\frac{-1}{2h^2}\|Q_j J_1 - q_{1,k}\|^2)}{\sum_k \exp(\frac{-1}{2h^2}\|Q_j J_1 - q_{1,k}\|^2)}$. This expression can be interpreted as follows. The image Z_1 is constructed from patches $\{z_{1,j}\}$. The patch $z_{1,j}$ is a weighted average of the K nearest neighbors of the jth patch in J_1, weighted according to their similarity to this patch. Here, M is the patch width.

The gradient of the data term is given by

$$\lambda \nabla_{J_1} E_{\text{data}} = \sum_k \frac{\delta_k \delta_k^T (J_1 - I_1^c)}{w_k^{1,data}}, \tag{28}$$

where

$$w_k^{1,data} = \frac{1}{\lambda}\sqrt{(\delta_k^T(I_1^c - J_1))^2 + \epsilon^2}. \tag{29}$$

Finally, the gradient of the correspondence term is given by

$$\lambda \nabla_{J_1} E_{\text{corr}} = \sum_k \frac{C_k \delta_k \delta_k^T (J_1 - J_{2\to 1})}{m_k^{1,data}} \tag{30}$$

where

$$m_k^{1,data} = \frac{1}{\lambda} \sqrt{(\delta_k^T (J_{2\to 1} - J_1))^2 + \epsilon^2}. \tag{31}$$

Setting to zero the sum of the gradients of these three energy terms, and noting that up to boundary effects, $\frac{1}{M^2}\sum_j Q_j^T Q_j$ equals the identity matrix, we obtain (in spatial coordinates)

$$\left(\tfrac{M^2}{h^2} + \frac{1}{W_{1,data}(x,y)} + \frac{C(x,y)}{M_{1,data}(x,y)} \right) J_1(x,y) = \quad \tfrac{M^2}{h^2} Z_1(x,y) + \frac{1}{W_{1,data}} I_1^c(x,y) + \frac{C(x,y)}{M_{1,data}(x,y)} J_{2\to 1}(x,y). \tag{32}$$

Similarly, for the second view,

$$\left(\tfrac{M^2}{h^2} + \frac{1}{W_{2,data}(x,y)} + \frac{C(x,y)}{M_{2,data}(x,y)} \right) J_2(x,y) = \tfrac{M^2}{h^2} Z_2(x,y) + \frac{1}{W_{2,data}} I_2^c(x,y) + \frac{C(x,y)}{M_{2,data}(x,y)} J_{1\to 2}(x,y). \tag{33}$$

4. APPENDIX D: Derivation of the transformation update step under the appearance consistency loss

To minimize the objective function

$$
\begin{aligned}
&E(u_1, v_1, u_2, v_2) = \\
&\lambda \iint \psi(\|\mathcal{T}_1^{-1} J_1(x,y) - I_1(x,y)\|^2) + \psi(\|\mathcal{T}_2^{-1} J_1(x,y) - I_2(x,y)\|^2) + \psi(\|\mathcal{T}_1^{-1} J_1(x,y) - \mathcal{T}_2^{-1} J_2(x,y)\|^2) dxdy \\
&\qquad + \alpha_r \iint \psi(\|\nabla u_1(x,y)\|^2 + \|\nabla v_1(x,y)\|^2) dxdy + \alpha_c \iint \psi(\|\nabla w_x(x,y)\|^2 + \|\nabla w_y(x,y)\|^2) dxdy \quad (34)
\end{aligned}
$$

with respect to $\mathcal{T}_1$ and $\mathcal{T}_2$ we set the gradient with respect to $\mathcal{T}_1$ and $\mathcal{T}_2$ to zero. We follow Liu's [3] IRLS approach to simultaneously update the two flow fields.

References

[1] C. S. Burrus. Iterative re-weighted least-squares. *OpenStax-CNX Web site. http://cnx. org/content/m45285/1.12*, 2012. 2

[2] T. Dekel, T. Michaeli, M. Irani, and W. T. Freeman. Revealing and modifying non-local variations in a single image. *ACM Transactions on Graphics (TOG)*, 34(6):227, 2015. 4

[3] C. Liu. *Beyond pixels: exploring new representations and applications for motion analysis*. PhD thesis, Massachusetts Institute of Technology, 2009. 1, 3, 6