

# Perceptually Motivated Benchmark for Video Matting

Mikhail Erofeev<sup>1</sup>  
merofeev@graphics.cs.msu.ru

Yury Gitman<sup>1</sup>  
ygitman@graphics.cs.msu.ru

Dmitriy Vatolin<sup>1</sup>  
dmitriy@graphics.cs.msu.ru

Alexey Fedorov<sup>1</sup>  
afedorov@graphics.cs.msu.ru

Jue Wang<sup>2</sup>  
juewang@adobe.com

<sup>1</sup> Lomonosov Moscow State University  
Moscow, Russia

<sup>2</sup> Adobe Systems  
Seattle, WA, United States

---

## Abstract

Despite recent progress in the field of video matting, neither public data sets nor even a generally accepted method of measuring quality has yet emerged. In this paper we present an online benchmark for video-matting methods. Using chroma keying and a reflection-aware stop-motion capturing procedure, we prepared 12 test sequences. Then, using subjective data, we performed extensive comparative analysis of different quality metrics. The goal of our benchmark is to enable better understanding of current progress in the field of video matting and to aid in developing new methods.

## 1 Introduction

Formally, matting is an inverse alpha-compositing problem: i.e., given pixel  $I$ , we want to find transparency value  $\alpha \in [0; 1]$ , foreground pixel  $F$ , and background pixel  $B$  so that

$$I = \alpha F + (1 - \alpha)B. \quad (1)$$

The problem is ill posed yet solvable by considering the affinity of pixels in natural images. Matting of natural images is well studied [5, 6, 8, 9, 10, 12, 13, 14, 18, 23, 24, 26], and according to [15], natural-image matting algorithms are continuously improving.

Video matting is a relatively new research direction that arose recently as available processing power increased. Applied to video, matting has two special requirements: tolerance of sparse user input and temporal coherence of the resulting transparency values. Developers of new methods achieve both requirements either explicitly by subjecting transparency values to temporal-smoothness constraints or implicitly by propagating the inner parameters required to solve for alpha.

Despite the rising interest, research in the field of video matting is still weakly organized. In fact, many developers estimate the quality of their methods by visual comparison [2, 8, 21].

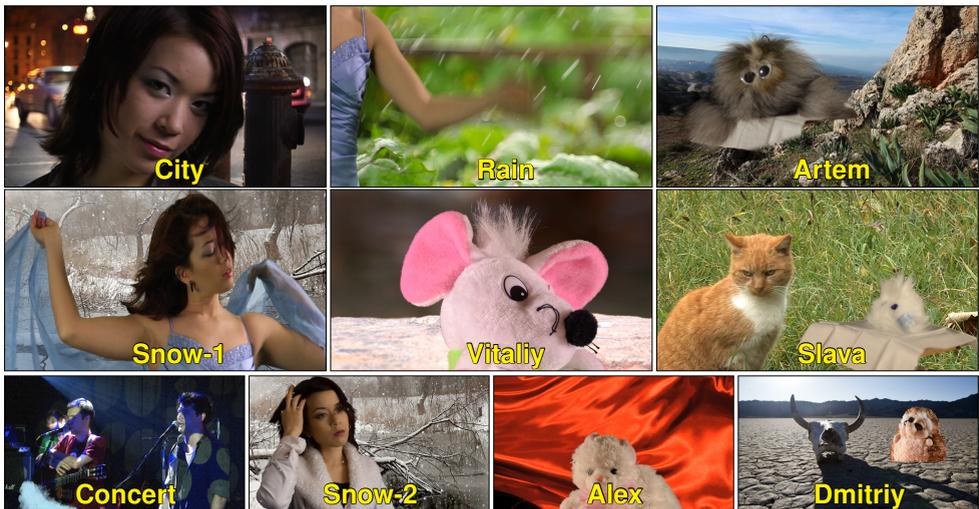


Figure 1: Preview frames of our test sequences.

The two main challenges facing an effective comparison are preparation of the data set and choice of a quality metric. In this paper we address both challenges and describe a benchmark, available at [videomattng.com](http://videomattng.com), that provides a comparison, two training sequences with ground-truth transparency, and multiple visualizations for convenient analysis of the comparison results.

To prepare the data set, we imposed four requirements on our test sequences: high quality for the ground-truth transparency, natural appearance, complexity, and diversity. To satisfy the first two requirements, we used two different techniques of foreground-object capture: namely, capture in front of a green screen and sequential photography against different backgrounds. We composed the extracted objects over a set of challenging backgrounds (see Figure 1) and prepared several trimap-width gradations.

Having obtained our test sequences, we conducted extensive subjective comparison of 12 matting methods. Using the collected data, we then compared different quality metrics. The results showed that alpha temporal coherence is significantly more important to human perception of video-matting quality than accuracy and that despite the imperfection of optical-flow techniques, use of these techniques can improve estimation of matte temporal consistency. The importance of temporal consistency, however, declines with increasing motion speed.

## 2 Related work

Although in this paper we introduce the very first benchmark for video matting, developers of new methods have long conducted simple comparisons to prove the superiority of their approaches. For instance, Lee et al. [10] proposed generalizations of robust matting [13] (both sampling and smoothing parts) to video sequences. They compared the performance of their method against frame-by-frame robust matting by computing the mean square error (MSE) of a single rendered sequence. Additionally, they measured temporal coherence as the ratio of temporal derivatives (RoTD) for the estimated alpha and the initial sequence.

Tang et al. [22] proposed generalization of closed-form matting [13] and compared their

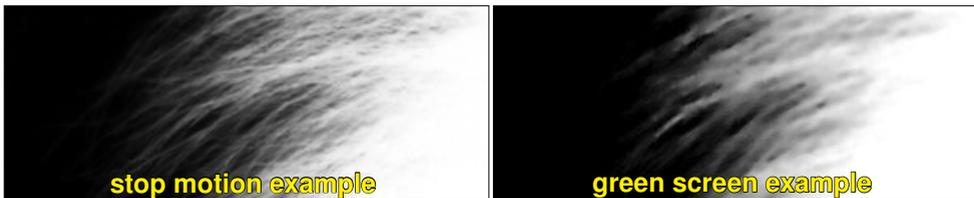


Figure 2: Alpha mattes from chroma keying and stop-motion capture for the same image region. The stop-motion result is significantly better at preserving details.

method against a frame-by-frame version and several video-segmentation methods—e.g., [10]. They used a set of blue-screen videos, and to estimate performance they used the mean absolute error (MAE) and its first derivative.

Brosch et al. [9] applied the recently proposed cost-volume filtering [16] to video matting and compared their method against a per-frame prototype using non-reference RoTD.

Li et al. [14] created 3D KNN matting [5] and compared it against a multiframe nonlocal matting Laplacian [6] using RoTD.

One of the most comprehensive comparisons appears in the work of Schahrian et al. [19]. The authors extended their previously proposed method of weighted color and texture matting [17] to the case of video and compared it with [10, 9] using the MSE between alpha values and the MSE between their temporal derivatives on a set of five green-screen videos.

In summary, most researchers perform only basic comparisons against a few competitors. Often, for the sake of simplicity, they sacrifice estimation of per-frame accuracy and measure only temporal stability using the non-reference RoTD metric [9, 14]. In some cases, researchers only employ a visual comparison [9, 6, 19]. There are thus several approaches to estimating the quality of different methods but so far no way to compare the performance of these approaches.

### 3 Data set

The crucial part of our proposed benchmark is a set of test sequences and corresponding ground-truth transparencies. The data set consists of five moving objects captured in front of a green plate and seven captured using the stop-motion procedure described below. We composed the objects over a set of background videos with various levels of 3D camera motion, color balance, and noise. We published ground-truth data for two stop-motion sequences and hid the rest to ensure fairness for the online benchmark that we introduce in Section 4.4.

Using thresholding and morphological operations on ground-truth alpha mattes, we generated trimaps, i.e. maps indicating if pixel certainly belongs to foreground, background or its transparency value needs to be solved. By varying width of unknown region we generated three trimap-width gradations: narrow, medium, and wide.

#### 3.1 Chroma keying

Chroma keying is a common practice in the film industry: the cinematographer captures an actor in front of a green or blue screen, then the VFX expert replaces the background using special software.

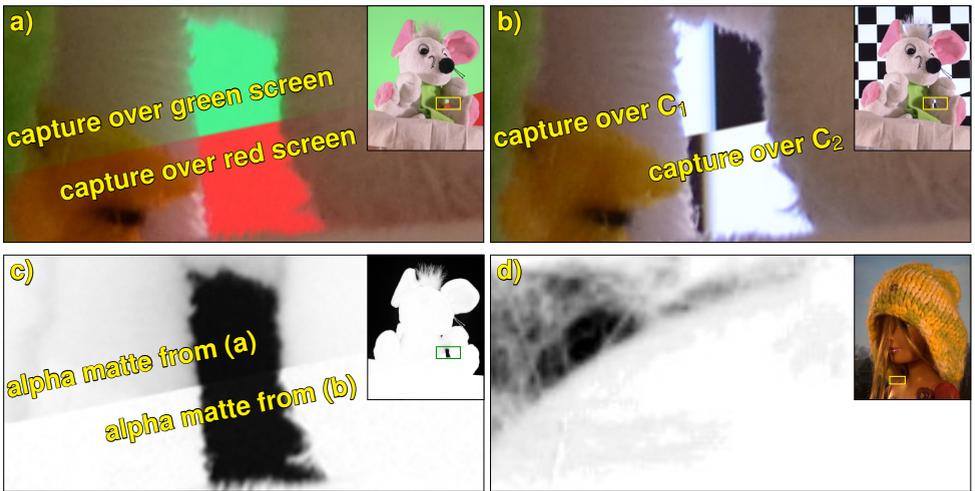


Figure 3: The problem of screen reflection and a proposed solution. **a)** Plush toy reflects color screen near the edges. **b)** Different checkerboard screens share the same white reflection. **c)** Alpha mattes computed using shots against solid color and checkerboard screens. **d)** Example of reflection problems in ground-truth image from [alphamattings.com](http://alphamattings.com).

For our research we used five green-screen video sequences with a significant amount of semitransparency—e.g., hair, motion blur, and smoke—provided by Hollywood camera work [47]. We extracted alpha mattes and corresponding foregrounds using The Foundry Keylight [25]. To improve quality for nonuniformly shaded green screens, we used multiple nonoverlapping instances of Keylight and adjusted each instance to deliver the best performance in its local region.

Chroma keying enabled us to obtain alpha mattes of natural-looking objects with arbitrary motion. Nevertheless, this technique cannot guarantee that the alpha maps are natural, because it assumes the screen color is absent from the foreground object (see Figure 2). To get alpha maps that have a very natural appearance, we used the stop-motion method.

### 3.2 Stop-motion

We designed the following procedure to perform stop-motion capture: an object with a fuzzy edge sits on the platform in front of an LCD monitor. The object rotates in small, discrete steps along a predefined 3D trajectory, controlled by two servomotors connected to a computer. After each step, the digital camera in front of the setup captures the motionless object against a set of background images. At the end of the process, we remove the object, and the camera again captures all of the background images.

Following [45], we can solve for transparency values in a system of alpha-compositing equations (see Equation 1). Instead, however, we use an extended system to allow for lighting variations caused by background-image changes (see Figure 3a):

$$\begin{cases} I_1 = \alpha F + (1-\alpha)B_1 + h * B_1 \\ \dots \\ I_n = \alpha F + (1-\alpha)B_n + h * B_n, \end{cases} \quad (2)$$

where  $I_1, I_2, \dots, I_n$  are images of the object over different backgrounds  $B_1, B_2, \dots, B_n$ ;  $*$  denotes convolution;  $h$  is a set of linear filters collecting light rays from the screen according to the object’s reflectance and Lambert’s cosine law; and  $\alpha$  and  $\alpha F$  are, respectively, the unknown transparency map and the object composite over a solid black background. By using the extended system, we avoid false transparency in a reflecting object’s area (see Figures 3c and 3d).

Because of the reflectance term, System of equations 2 is underdetermined, yet we are free to change the feasible solutions by choosing  $B$ . For instance, solid-color backgrounds satisfy  $h * B_i = (h * 1)B_i$  and give us an unambiguous  $\alpha F$  value. There are probably no such background images that uniquely determine  $\alpha$ , so we need to make an additional assumption about  $h$ . We can use black-and-white checkerboard images  $C_1$  and  $C_2 = 1 - C_1$  with small cells; in addition, considering that  $h$  is a low-pass filter and assuming that  $h * C_1 \approx h * C_2$  (see Figure 3b), we derive an explicit expression for the transparency values:

$$\alpha \approx 1 - \frac{(I_1 - I_2, C_1 - C_2)}{\|C_1 - C_2\|^2}. \quad (3)$$

Equation 3 enables us to compute  $\alpha$  everywhere except for small bands around pattern edges where the difference  $\|C_1 - C_2\|$  is small because of the optical blur. To eliminate this uncertainty we shift the pattern in horizontal, vertical, and diagonal directions and capture the object against the shifted patterns and their inversions. Thus, for each pixel we have at least one system with a significantly nonzero value of  $\|C_1 - C_2\|$ .

## 4 Objective comparison

### 4.1 Perceptual data

After obtaining the data set, the next important part of the benchmark process is quality measurement. For video matting, the general approach is to combine estimates of overall method accuracy, typically measured using mean square error (MSE) and temporal-coherence error [14, 19, 22]. MSE is still the best known measure of overall accuracy, yet it provides only 30% to 50% correlation with human perception [15]. Measurement of temporal consistency is even less well studied, and approaches vary among authors.

To support our choice of quality metric, we collected subjective pairwise comparisons of 12 matting methods applied to the videos from our data set. In particular, we showed participants a sequence of video pairs; for each pair, we asked them to choose the video with better quality or to indicate that the videos are approximately equal. For the sake of contrast, we composed the results of the methods over blue and yellow checkerboard images, and since many pairs differ only during brief intervals, we divided long sequences into shorter segments of 50 to 60 frames. Among our viewers were 101 volunteers and 442 paid participants from Amazon Mechanical Turk. We offered paid participants \$0.05 for 23 pairs, 3 of which were hidden quality-control comparisons between ground truth and a low-quality method; to accept the data from a given individual, we required correct choices for all control comparisons (we consider only one session per participant). In total, we collected 12,629 comparisons (3,789 from volunteers and 8,840 from paid participants), equally distributed among methods and sequences. All collected data is available at [videomattng.com/subj](http://videomattng.com/subj).

Using the Bradley-Terry model [9], we transformed the data into subjective scores (see Figure 4a). As expected, the best score pertains to the Adobe After Effects Refine Edge tool [23].

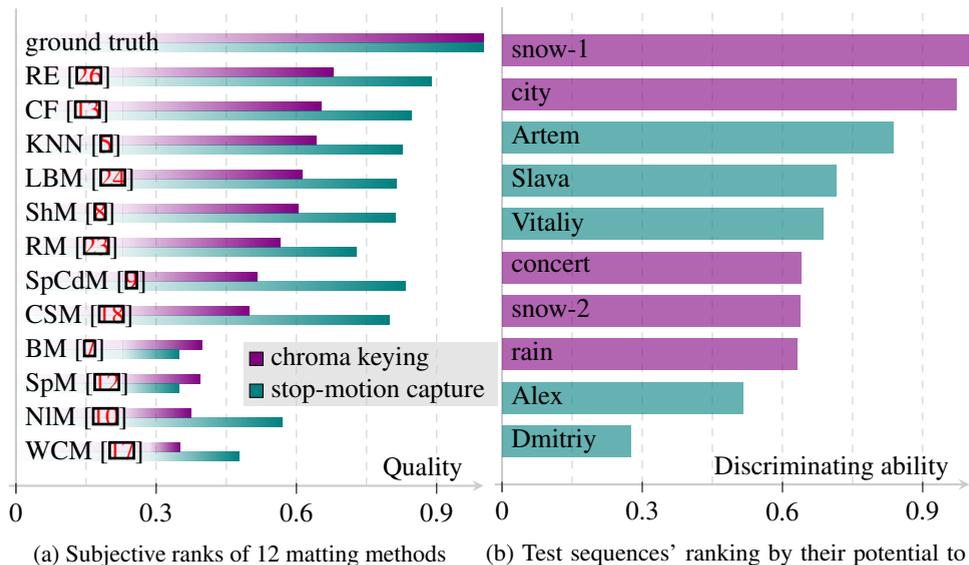


Figure 4: Results of pairwise subjective comparison of matting methods.

More interestingly, the matting Laplacian first proposed in 2006 [13] remains one of the best methods, probably because of its stability with respect to changes in the input image. Another reason may be that many researchers have overfit their methods to the images in [13].

In addition, the collected data helps us to estimate the quality of our test sequences. Since different test sequences have different potential to distinguish weak and strong methods, researchers may prefer to check their new methods against sequences that are better able to discriminate among them. Specifically, we define the discriminating ability of a given sequence as the average of all inter-comparisons among the five best and five worst methods taken from all the data (we assign a value of +1 for selection of the first method, -1 for selection of the second method, and 0 indicating that the methods were deemed approximately equal). Figure 4b shows the discriminating capabilities for all of our test sequences. Predictably, sequences captured in front of a green screen are more difficult and thus are better for discriminating among methods.

## 4.2 Quality metric

Even given ground-truth data, there is no generally accepted way of estimating video-matting quality. To support our choice of quality metric, we compared 12 different candidates by their correlation with subjective data. Roughly, our candidates fall into three subsets defined by similar behavior: two metrics for matte accuracy, eight metrics for matte temporal coherence, and two motion-aware metrics for temporal coherence.

**Accuracy metrics** process frames independently and thus do not take into account any interframe impairments. In this study, we avoid considering complex metrics based on modeling of the natural-image alpha matte; our accuracy metrics independently process each pixel and therefore capture only mean per-frame difference between the evaluated and ground-truth sequences.

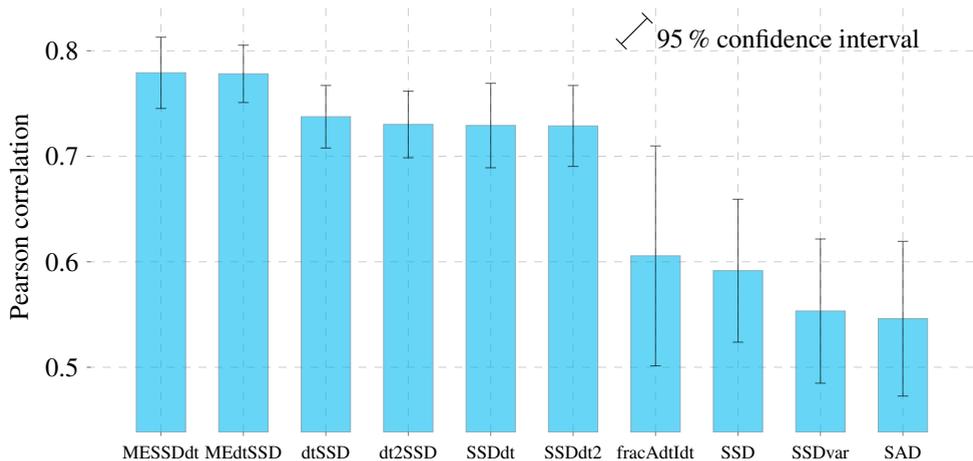


Figure 5: Comparison of objective quality metrics according to correlation of their values with perceptual data.

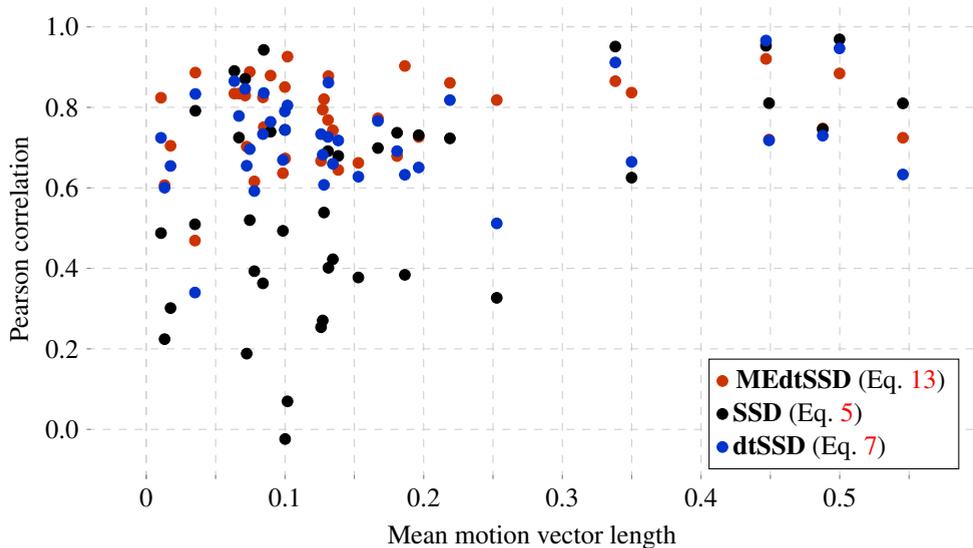
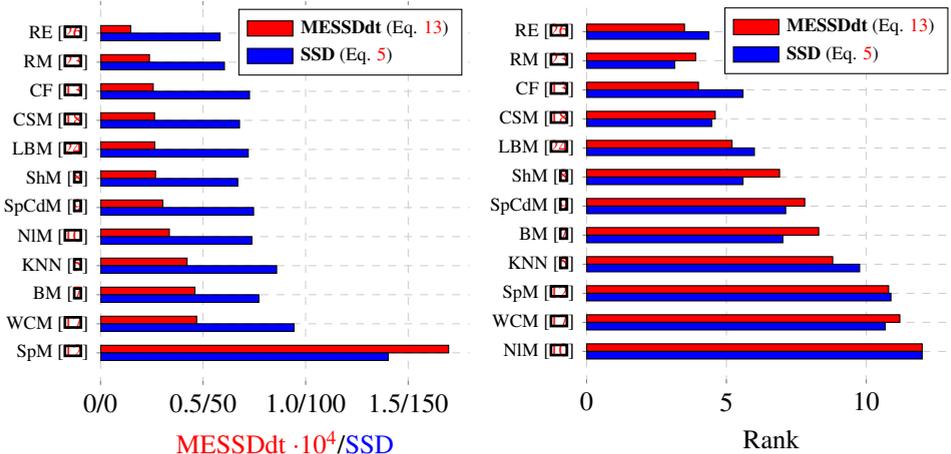


Figure 6: Influence of average motion speed on performance of spatial, temporal, and temporal motion-aware metrics.



(a) Comparison results for matting methods applied to the "City" test sequence using temporal-coherence and spatial-error metrics

(b) Mean method ranks for all test sequences

Figure 7: Objective comparison results.

Let  $\#$  denote the number of frames. Also, for pixel  $p$  of frame  $t$ , let  $\alpha_{p,t}$  denote the transparency value of the video matting under consideration and let  $\alpha_{p,t}^{GT}$  denote the ground truth. Then,

$$SAD = \frac{1}{\#} \sum_{t,p} |\alpha_{p,t} - \alpha_{p,t}^{GT}|, \quad (4)$$

$$SSD = \frac{1}{\#} \sum_t \sqrt{\sum_p (\alpha_{p,t} - \alpha_{p,t}^{GT})^2}. \quad (5)$$

**Temporal-coherence metrics** capture unexpected alpha temporal changes and ignore temporally coherent errors. Similarly to the case of accuracy metrics, we make no assumptions about frame structure. Computation of two metrics, however, includes estimation of optical flow and thus inherits the assumptions of optical flow. We included these metrics in our comparison despite their dependency on an imperfect optical-flow implementation, since even in the case of false motion vectors we expect their performance to be similar to that of motion-unaware metrics (every motion-unaware metric can be interpreted as a motion-aware metric based on a constant zero-valued field of motion vectors). In the case of vectors that closely represent the true motion, the performance of these metric should be much higher.

For all experiments in this section, we use a block-based motion-estimation algorithm [24] computed for a ground-truth sequence composed over solid-green background. Note that as a result, motion-aware metrics avoid giving an unfair advantage to matting methods based on a similar motion-estimation approach, since these methods lack a ground-truth sequence.

Using the same notation for  $\#$ ,  $p$ ,  $t$ ,  $\alpha_{p,t}$ , and  $\alpha_{p,t}^{GT}$  and denoting motion vectors by  $v_p$ , we define

$$\mathbf{SSDdt} = \frac{1}{\#} \sum_t \left| \frac{d \sum_p (\alpha_{p,t} - \alpha_{p,t}^{GT})^2}{dt} \right|, \quad (6) \quad \mathbf{dtSSD} = \mathbf{SSD} \left( \frac{d\alpha_{p,t}}{dt}, \frac{d\alpha_{p,t}^{GT}}{dt} \right), \quad (7)$$

$$\mathbf{SSDdt2} = \frac{1}{\#} \sum_t \left| \frac{d^2 \sum_p (\alpha_{p,t} - \alpha_{p,t}^{GT})^2}{dt^2} \right|, \quad (8) \quad \mathbf{dt2SSD} = \mathbf{SSD} \left( \frac{d^2 \alpha_{p,t}}{dt^2}, \frac{d^2 \alpha_{p,t}^{GT}}{dt^2} \right), \quad (9)$$

$$\mathbf{SSDvar} = \text{var}_t \left( \sum_p (\alpha_{p,t} - \alpha_{p,t}^{GT})^2 \right), \quad (10) \quad \mathbf{AdtIdt} = \frac{1}{\#} \sum_{t,p} \left| \frac{d\alpha_{p,t}}{dt} \right| / \left( 1 + \left| \frac{dv_p}{dt} \right| \right), \quad (11)$$

$$\mathbf{MESSDdt} = \frac{1}{\#} \sum_{t,p} \left| (\alpha_{p,t} - \alpha_{p,t}^{GT})^2 - (\alpha_{p+v_p,t+1} - \alpha_{p+v_p,t+1}^{GT})^2 \right|, \quad (12)$$

$$\mathbf{MEdtSSD} = \frac{1}{\#} \sum_{t,p} \left( \left| \alpha_{\alpha,t} - \alpha_{p+v_p,t+1} \right| - \left| \alpha_{p,t}^{GT} - \alpha_{p+v_p,t+1}^{GT} \right| \right)^2. \quad (13)$$

Authors of [9, 10, 14] used non-reference metric **AdtIdt** for temporal coherence evaluation. Shahrian et al. used **dtSSD**. Tang et al. used metric similar to **SSDdt**. Finally, **SSDdt2**, **dtSSD**, **dt2SSD** and **SSDvar** are slight variations of above mentioned metrics. Motion-aware metrics **MESSDdt** and **MEdtSSD** are straightforward generalizations of **SSDdt** and **dtSSD** respectively with use of optical flow.

For each metric, we estimate scores for the set of 12 video-matting methods applied to 10 test sequences. Then, for the logarithm of each score we compute the Pearson correlation coefficient with the subjective data. Figure 5 shows the mean correlation values. The results of our comparison revealed greater perceptual importance of temporal coherence than accuracy. As expected, taking motion into account increases mean correlation even more. But the relative inferiority of accuracy metrics emerges mostly for slow-motion sequences; it becomes almost insignificant as motion speed increases. One possible explanation is that the human visual system pays less attention to temporal inconsistencies. Figure 6 shows the exact dependency.

### 4.3 Evaluation results

Since we believe both accuracy and temporal coherence are important indicators of video-matting quality, we independently compared matting methods using **SSD** and **MESSDdt**. At the time of publication, our benchmark offers comparison of 12 matting methods over 10 test sequences. We show results of the comparison for the ‘‘City’’ test sequence in Figure 7a and show average method ranks for all of our test sequences in Figure 7b.

Additionally, Figure 8 shows **MESSDdt** values for various gradations of trimap’s unknown area width. The figure clearly demonstrates that methods based on color sampling [10, 8, 13] are less robust with respect to increasing trimap width than alpha-propagation methods [10, 13, 14] are.

Currently, all methods in our comparison require a per-frame trimap, but we encourage researchers to submit methods that require a trimap only for the first and last frames, since support for sparse user input is very important to video matting.

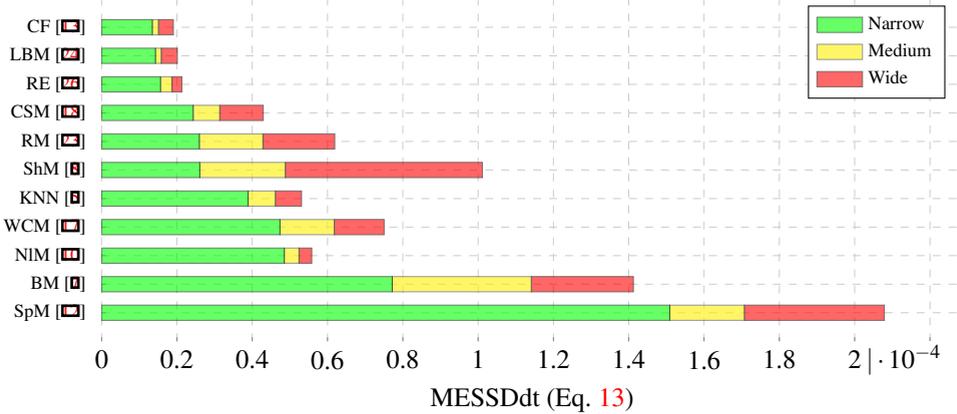


Figure 8: Performance of methods for various widths of trimap’s unknown area on “Artem” sequence.

## 4.4 Website

To simplify access to the benchmark and enable addition of new methods, we created the [videomattng.com](http://videomattng.com) website, which contains scatterplots and rating tables for different quality metrics. In addition, results for participating methods are available for viewing on a player equipped with a movable zoom region. Besides the comparison, the website includes instructions for new participants, links to test videos, and two sequences with ground-truth transparency.

## 5 Conclusion

In this paper we presented an online benchmark, available at [videomattng.com](http://videomattng.com), that allows researchers to evaluate their video-matting methods. We designed a procedure for obtaining high-quality ground-truth transparencies and performed extensive analysis of a possible quality metric based on subjective data from human perception.

The reported study was funded by RFBR according to the research project 15-01-08632 A.

## References

- [1] Xue Bai, Jue Wang, David Simons, and Guillermo Sapiro. Video snapcut: Robust video object cutout using localized classifiers. *ACM Transactions on Graphics (TOG)*, 28(3): 70:1–70:11, 2009. doi: 10.1145/1531326.1531376.
- [2] Xue Bai, Jue Wang, and David Simons. Towards temporally-coherent video matting. In *International Conference on Computer Vision (ICCV)*, pages 63–74, 2011. doi: 10.1007/978-3-642-24136-9\_6.
- [3] Ralph Allan Bradley and Milton E Terry. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, pages 324–345, 1952.

- [4] Nicole Brosch, Asmaa Hosni, Christoph Rhemann, and Margrit Gelautz. Spatio-temporally coherent interactive video object segmentation via efficient filtering. In *Pattern Recognition*, volume 7476, pages 418–427, 2012. doi: [10.1007/978-3-642-32717-9\\_42](https://doi.org/10.1007/978-3-642-32717-9_42).
- [5] Qifeng Chen, Dingzeyu Li, and Chi-Keung Tang. KNN matting. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 35(9):2175–2188, 2013. doi: [10.1109/TPAMI.2013.18](https://doi.org/10.1109/TPAMI.2013.18).
- [6] Inchang Choi, Minhaeng Lee, and Yu-Wing Tai. Video matting using multi-frame nonlocal matting laplacian. In *European Conference on Computer Vision (ECCV)*, pages 540–553, 2012. doi: [10.1007/978-3-642-33783-3\\_39](https://doi.org/10.1007/978-3-642-33783-3_39).
- [7] Yung-Yu Chuang, Brian Curless, David H. Salesin, and Richard Szeliski. A bayesian approach to digital matting. In *Computer Vision Pattern Recognition (CVPR)*, volume 2, pages II–264–II–271, 2001. doi: [10.1109/CVPR.2001.990970](https://doi.org/10.1109/CVPR.2001.990970).
- [8] Eduardo S.L. Gastal and Manuel M. Oliveira. Shared sampling for real-time alpha matting. *Computer Graphics Forum*, 29(2):575–584, 2010. doi: [10.1111/j.1467-8659.2009.01627.x](https://doi.org/10.1111/j.1467-8659.2009.01627.x).
- [9] Jubin Johnson, Deepu Rajan, and Hisham Cholakkal. Sparse codes as alpha matte. In *Computer Graphics Forum*, volume 32, pages 245–253, 2013.
- [10] Philip Lee and Ying Wu. Nonlocal matting. In *Computer Vision Pattern Recognition (CVPR)*, pages 2193–2200, 2011. doi: [10.1109/CVPR.2011.5995665](https://doi.org/10.1109/CVPR.2011.5995665).
- [11] Sun-Young Lee, Jong-Chul Yoon, and In-Kwon Lee. Temporally coherent video matting. *Graphical Models*, 72(3):25–33, 2010. doi: [10.1016/j.gmod.2010.03.001](https://doi.org/10.1016/j.gmod.2010.03.001).
- [12] A. Levin, A. Rav Acha, and D. Lischinski. Spectral matting. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 30(10):1699–1712, 2008. doi: [10.1109/TPAMI.2008.168](https://doi.org/10.1109/TPAMI.2008.168).
- [13] Anat Levin, Dani Lischinski, and Yair Weiss. A closed-form solution to natural image matting. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 30(2):228–242, 2008. doi: [10.1109/TPAMI.2007.1177](https://doi.org/10.1109/TPAMI.2007.1177).
- [14] Dingzeyu Li, Qifeng Chen, and Chi-Keung Tang. Motion-aware KNN Laplacian for video matting. In *International Conference on Computer Vision (ICCV)*, pages 3599–3606, 2013. doi: [10.1109/ICCV.2013.447](https://doi.org/10.1109/ICCV.2013.447).
- [15] Christoph Rhemann, Carsten Rother, Jue Wang, Margrit Gelautz, Pushmeet Kohli, and Pamela Rott. A perceptually motivated online benchmark for image matting. In *Computer Vision Pattern Recognition (CVPR)*, pages 1826–1833, 2009. [www.alphamatting.com](http://www.alphamatting.com).
- [16] Christoph Rhemann, Asmaa Hosni, Michael Bleyer, Carsten Rother, and Margrit Gelautz. Fast cost-volume filtering for visual correspondence and beyond. In *Computer Vision Pattern Recognition (CVPR)*, pages 3017–3024, 2011. doi: [10.1109/CVPR.2011.5995372](https://doi.org/10.1109/CVPR.2011.5995372).

- [17] E. Shahrian and D. Rajan. Weighted color and texture sample selection for image matting. In *Computer Vision Pattern Recognition (CVPR)*, pages 718–725, 2012. doi: [10.1109/CVPR.2012.6247741](https://doi.org/10.1109/CVPR.2012.6247741).
- [18] E. Shahrian, D. Rajan, B. Price, and S. Cohen. Improving image matting using comprehensive sampling sets. In *Computer Vision Pattern Recognition (CVPR)*, pages 636–643, 2013. doi: [10.1109/CVPR.2013.88](https://doi.org/10.1109/CVPR.2013.88).
- [19] Ehsan Shahrian, Brian Price, Scott Cohen, and Deepu Rajan. Temporally coherent and spatially accurate video matting. *Computer Graphics Forum*, 33(2):381–390, 2014. doi: [10.1111/cgf.12297](https://doi.org/10.1111/cgf.12297).
- [20] Karen Simonyan, Sergey Grishin, Dmitriy Vatolin, and Dmitriy Popov. Fast video super-resolution via classification. In *International Conference on Image Processing (ICIP)*, pages 349–352, 2008. doi: [10.1109/ICIP.2008.4711763](https://doi.org/10.1109/ICIP.2008.4711763).
- [21] Mikhail Sindeev, Anton Konushin, and Carsten Rother. Alpha-flow for video matting. In *Asian Conference on Computer Vision (ACCV)*, pages 438–452, 2013. doi: [10.1007/978-3-642-37431-9\\_34](https://doi.org/10.1007/978-3-642-37431-9_34).
- [22] Zhen Tang, Zhenjiang Miao, Yanli Wan, and Dianyong Zhang. Video matting via opacity propagation. *The Visual Computer*, 28(1):47–61, 2012. doi: [10.1007/s00371-011-0598-3](https://doi.org/10.1007/s00371-011-0598-3).
- [23] Jue Wang and Michael F. Cohen. Optimized color sampling for robust matting. In *Computer Vision Pattern Recognition (CVPR)*, pages 1–8, 2007. doi: [10.1109/CVPR.2007.383006](https://doi.org/10.1109/CVPR.2007.383006).
- [24] Yuanjie Zheng and C. Kambhamettu. Learning based digital matting. In *International Conference on Computer Vision (ICCV)*, pages 889–896, 2009. doi: [10.1109/ICCV.2009.5459326](https://doi.org/10.1109/ICCV.2009.5459326).
- [25] Keylight. <http://www.thefoundry.co.uk/products/plugins/keylight/>, 2015.
- [26] AE refine edge tool. <http://www.adobe.com/en/products/aftereffects.html>, 2015.
- [27] Hollywood camera work. <http://www.hollywoodcamerawork.com/greenscreenplates.html>, 2015.