Photo2ClipArt: Image Abstraction and Vectorization Using Layered Linear Gradients
Jean-Dominique Favreau, Florent Lafarge, Adrien Bousseau Bousseau

To cite this version:

HAL Id: hal-01581981
https://inria.hal.science/hal-01581981
Submitted on 5 Sep 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Photo2ClipArt: Image Abstraction and Vectorization Using Layered Linear Gradients

JEAN-DOMINIQUE FAVREAU, FLORENT LAFARGE, ADRIEN BOUSSEAU, Inria - Université Côte d’Azur

Fig. 1. Given a segmented bitmap image as input (a), our method generates an abstract, layered vector clipart, where each layer is filled with an opaque or semi-transparent linear color gradient (b). By expressing the image as a stack of linear color gradients, our vector graphics reproduce the visual style of traditional cliparts and are easy to edit (c). In this example, we turned the lady bug blue (top) and replaced its dots by little stars (bottom). The black dots in the segmentation indicate opaque regions selected by the user to initialize or constrain our algorithm.

We present a method to create vector cliparts from photographs. Our approach aims at reproducing two key properties of cliparts: they should be easily editable, and they should represent image content in a clean, simplified way. We observe that vector artists satisfy both of these properties by modeling cliparts with linear color gradients, which have a small number of parameters and approximate well smooth color variations. In addition, skilled artists produce intricate yet editable artworks by stacking multiple gradients using opaque and semi-transparent layers. Motivated by these observations, our goal is to decompose a bitmap photograph into a stack of layers, each layer containing a vector path filled with a linear color gradient. We cast this problem as an optimization that jointly assigns each pixel to one or more layer and finds the gradient parameters of each layer that best reproduce the input. Since a trivial solution would consist in assigning each pixel to a different, opaque layer, we complement our objective with a simplicity term that favors decompositions made of few, semi-transparent layers. However, this formulation results in a complex combinatorial problem combining discrete unknowns (the pixel assignments) and continuous unknowns (the layer parameters). We propose a Monte Carlo Tree Search algorithm that efficiently explores this solution space by leveraging layering cues at image junctions. We demonstrate the effectiveness of our method by reverse-engineering existing cliparts and by creating original cliparts from studio photographs.

This is the author’s version of the work. It is posted here for your personal use. Not for redistribution. The definitive version was published in ACM Transactions on Graphics, https://doi.org/10.1145/3130800.3130888.

CCS Concepts: • Computing methodologies → Non-photorealistic rendering;

Additional Key Words and Phrases: image abstraction, image stylization, vector graphics, vectorization, layers, transparency, color gradient

1 INTRODUCTION

Vector cliparts are widely used in graphic design to represent images in a compact and editable manner. Cliparts are also greatly appreciated for the abstract, stylized look produced by their constituent parametric paths and color fills. Our goal in this paper is to help novice users create such stylized vector artworks from photographs.

Our approach is inspired by the observation that skilled artists produce rich yet abstract cliparts by stacking, or layering, few parametric vector paths. For example, the inset shows the clipart of a red soda can where the texture, shadows and highlights are all represented as semi-transparent layers filled with constant colors or color gradients. Changing the color of the bottom layer is sufficient to obtain a convincing blue can. Following a similar strategy, our approach seeks to decompose the input image into a small number of semi-transparent layers, each layer filled with a constant color or a two-color linear gradient. By favoring a simple interpretation of the input, our method jointly abstracts and vectorizes the image to produce a stylized, editable clipart.

We formulate our approach as a joint labeling and fitting problem, where we want to assign each pixel to one or more layers such

1 We rasterized all vector arts shown in the paper to ensure that transparency is well displayed in different pdf viewers. Please refer to supplemental materials for the original vector files.
that approximating the resulting layers with linear color gradients yields a good reconstruction of the input image. We balance this reconstruction error with a simplicity term that penalizes the use of too many layers. However, the resulting optimization involves both discrete unknowns (the pixel assignments) and continuous ones (the color gradient parameters). Finding the optimal solution of such a problem would require evaluating all possible pixel assignments, which is impractical. Our main technical contribution is a stochastic algorithm that explores this solution space efficiently.

Our algorithm builds on several key ideas to drastically decrease the number of configurations it needs to evaluate. Our first idea is to exploit image junctions, which provide strong local layering and transparency cues [Sayim and Cavanagh 2011]. In particular, we assume that a X-junction occurs when a semi-transparent layer runs across two other layers, which gives us a small number of possible configurations. Our second idea is to use the local decompositions obtained from X-junctions to initialize a region-growing approach, which progressively assigns new pixels to their best-fitting existing regions and regions that should contain the layers of another region. Our objective of favoring a simple vectorization is similar in spirit to the work by Favreau et al. [2016] who convert line drawings into a small yet accurate set of Bezier curves, and to the work by Jeschke et al. [2011] and Zhao et al. [2017] who optimize for diffusion curves with simple color and shape respectively. Our originality is to target layered color images, which greatly increases the dimensionality of the problem and requires us to design a custom exploration mechanism of the solution space.

Image abstraction. Researchers in non-photorealistic rendering have proposed a variety of image filters to abstract and stylize photographs. Popular methods include the use of scale-space filtering [DeCarlo and Santella 2002] and edge-aware filtering [Winnemöller et al. 2006] to remove low-contrast details. However, such image abstraction filters are usually applied independently of image vectorization. In contrast, we obtain abstract, stylized cliparts by expressing the image with a small number of parametric color gradients. A similar idea of restricting the image formation model to achieve abstraction has been explored by Gerstner et al. [2012], who produce pixel art by converting a photograph into a low-resolution image with a reduced color palette.

Image decomposition. While our goal is to produce vector graphics, our work is also related to layered decomposition of bitmaps. In particular, Tan and colleagues [2015; 2016] decompose artworks into semi-transparent layers to perform a range of editing operations. Their first method takes as input a time lapse video of the artwork being painted [Tan et al. 2015], and as such is not applicable in our context. In a follow-up work [Tan et al. 2016], they observe that the dominant colors of an image correspond to the vertices of the convex hull of the image in RGB space. This analysis allows them to decompose the image into a stack of smooth semi-transparent layers, each layer corresponding to one of the dominant colors with per-pixel transparency. Similarly, Aksoy et al. [2017] propose a soft segmentation algorithm that decomposes an image into layers with per-pixel transparency, although they associate each layer with an RGB normal distribution rather than a constant color. We adopt a very different approach by representing our layers with linear color gradients amenable to vectorization. Our more constrained representation allows us to express each layer with a handful of parameters and a binary mask rather than with one continuous unknown per pixel. Our energy formulation also allows us to explicitly minimize the complexity of the output and to automatically find the layer ordering.

Our problem is also related to other ill-posed image decompositions such as image matting [Smith and Blinn 1996], reflection separation [Levin et al. 2004] and intrinsic images [Bell et al. 2014]. However, a major difference between these methods and our work is that they aim at separating only two rather than multiple layers, being foreground and background or reflectance and shading.
We synthesize the output image where we favor a small number of layers, each represented by a constant color or a linear gradient.

**Depth ordering and transparency from junctions.** Perceptual studies emphasize the role of image junctions in the perception of occlusion and transparency [Metelli 1974; Sayim and Cavanagh 2011]. In particular, T-junctions provide strong cues of local ordering between opaque layers [Jia et al. 2012; Liu et al. 2013] while luminance and chrominance patterns at X-junctions have been used for extracting transparent layers from images [D’Zmura et al. 1997; Singh and Huang 2003]. We follow a similar approach to identify transparent and opaque layers at X-junctions. However, while prior work derives heuristic rules to reason about layer ordering and transmittance, we use an energy formulation to solve for the layer configuration that best reconstructs the input image. Our energy also favors simple interpretations, which allows us to deal with configurations where no X-junctions occur, such as when a transparent region is entirely surrounded by another region.

### 3 PROBLEM FORMULATION

Our goal is to estimate a multilayer vectorial representation of an input image \( I \), where each layer is composed of:

- A supporting domain \( D \) covering a subset of pixels from the input image,
- A color gradient function \( C(p) \) that associates an RGB color to each pixel \( p \in D \).
- An opacity gradient function \( A(p) \) that associates an opacity to each pixel \( p \in D \). We set \( A(p) = 0 \) when \( p \notin D_n \).

We synthesize the output image \( I_n \) from a \( n \)-layer representation by recursively \( \alpha \)-blending the ordered layers

\[
I_n(p) = A_n(p)C_n(p) + (1 - A_n(p))I_{n-1}(p)
\]

where \( C_n \) and \( A_n \) are the color gradient function and opacity gradient function of layer \( n \). Figure 2 illustrates this representation.

![Multilayer representation](image)

**Fig. 2.** Multilayer representation. Our goal is to represent the input image (a) as a stack of opaque and semi-transparent layers (b) such that \( \alpha \)-blending the layers using Eq. 1 reproduces well the input (c).

### 3.1 Energy formulation

We denote by \( x = (D_1, C_1, A_1, ..., D_n, C_n, A_n) \), an output vectorial representation composed of \( n \) layers. We define three criteria to measure the quality of a configuration \( x \):

- **Faithfulness to input data** - the reconstructed image \( I_n \) has to be similar to the input image \( I \).
- **Simplicity of opacity functions** - the number of layers should be minimized to yield a compact and editable representation.
- **Simplicity of opacity functions** - semi-transparent layers should be favored over opaque ones to reduce occlusion.

Based on these criteria, we formulate an energy \( U \) of the form

\[
U(x) = (1 - \lambda)U_{\text{fidelity}}(x) + \lambda U_{\text{simplicity}}(x)
\]

where \( U_{\text{fidelity}} \) measures the faithfulness to input data and \( U_{\text{simplicity}} \) accounts for the overall simplicity of the reconstruction. The parameter \( \lambda \in [0, 1] \) weights the two terms.

We define \( U_{\text{fidelity}} \) as the RGB error under \( L_2 \) norm between the input image \( I \) and the reconstructed image \( I_n \):

\[
U_{\text{fidelity}}(x) = \frac{1}{|I|} \sum_{p \in I} \|I(p) - I_n(p)\|_2^2.
\]

We express \( U_{\text{simplicity}} \) as

\[
U_{\text{simplicity}}(x) = \frac{1}{N} \sum_{l=1}^{n} w_l
\]

where \( N \) is the maximal number of layers and \( w_l \) penalizes opacity functions according to their simplicity. We set \( w_l = 1 \) for semi-transparent layers and \( w_l = \beta > 1 \) for opaque layers. Figure 2...

**Photo2ClipArt: Image Abstraction and Vectorization**

Using Layered Linear Gradients • 180:3

would require estimating the continuous variables for which facilitates subsequent editing such as changing the color of variables (the parameters of the linear gradient functions). While decompositions of a few regions, which we obtain by exploiting a simple example in Figure 3. Our exploration starts from initial assignments, which is impractical despite this upper bound.

The number of layers and their supporting domains) and continuous plex optimization problem which combines discrete variables (the parameters of the linear gradient functions). While the number of layers can be infinite, we can reasonably assume that there is at most as many layers as input regions. Yet, since a region can appear in at most all N layers, finding the global minimum would require estimating the continuous variables for \( 2^N \) layer assignments, which is impractical despite this upper bound.

We address this combinatorial challenge by devising a scalable exploration strategy of the solution space, which we illustrate on a simple example in Figure 3. Our exploration starts from initial decompositions of a few regions, which we obtain by exploiting transparency assumptions on X-junctions, or by asking the user to indicate one or more opaque regions. We then expand these decompositions to the regions adjacent to the ones already decomposed. Each new region can either be assigned to one or more existing layers, or to a new layer, for which the color gradient is estimated. We thus obtain a set of possible assignments of the new regions, each forming a decomposition that can be further expanded to adjacent regions. We can represent the ensemble of decompositions generated by this approach as a tree, where each node corresponds to an intermediate decomposition, which branches to multiple decompositions after an expansion step. The leafs of this tree correspond to all the \( 2^N \) possible decompositions, and our goal is to find a low-energy leaf without exploring the entire tree.

We propose a stochastic algorithm for fast exploration of the solution space, which can be seen as a form of Monte Carlo Tree Search (MCTS) [Browne et al. 2012]. The main idea of this method is to build the solution tree incrementally and asymmetrically using a balance between exploration of the space and exploitation of the already explored configurations. Figure 4 illustrates the four successive operations performed at each iteration of an MCTS method.
Algorithm 1 Exploration mechanism

Initialize child nodes of the root (sec 4.5)
repeat
  Node selection: select a leaf node using tree policy (sec 4.1)
  Expansion: generate child nodes (sec 4.2)
  Reward: compute energy of child nodes (sec 4.3)
  Back-propagation: update tree policy (sec 4.4)
until stopping criterion (sec 4.5)

4.1 Node selection

The tree policy seeks to favor the exploration of high-quality configurations. To do so, each child node \( t_j \) of a node \( t \) is associated to a parent-to-child probability that is proportional to its energy and the energy of its siblings

\[
P_{t \rightarrow t_j} = \frac{\exp(-U(x_{t_j})) \mathbb{1}_{(t_j \in t^*)}}{\sum_{t_j \in t^*} \exp(-U(x_{t_j}))}
\]

where \( t^* \) is the set of child nodes of \( t \) and \( \mathbb{1} \) is the indicator function. The algorithm selects a terminal node for expansion according to the product of parent-to-child probabilities between the root and the node, i.e. the overall probability of reaching this configuration.

We further accelerate the search by restricting \( t^* \) to the \( k \) children nodes with the lowest energies. The parameter \( k \) offers a balance between spread and accuracy of the algorithm. We set \( k \) to 2 in all our experiments.

4.2 Tree expansion

Once a candidate terminal node is selected, the tree expansion step generates its child nodes by adding one or more regions to the ones already visited. We first present a naive generator that simply adds one region with all its possible layer assignments. We then detail a more efficient generator that exploits a transparency assumption on X-junctions to add multiple regions at once. We detect X-junctions as cliques of order 4 in the adjacency graph of the regions.

**Single-region expansion.** This generator inserts an unvisited region to the decomposition associated with the selected node. This region is chosen randomly among the regions adjacent to the ones already present in the decomposition. We assume that the new region can be part of any layer of its adjacent regions in the decomposition, as well as part of a new layer. Note that we do not consider layers of non-adjacent regions, which drastically reduces the number of child nodes while ensuring that each layer contains a single connected component by construction. Given the \( M \) layers from the adjacent regions, a naive enumeration gives \( 2^{M+1} \) child nodes. However, since opaque layers occlude all layers below them, we can discard all but one of the configurations that have the same visible layers, which further reduces the number of child nodes.

**X-junction expansion.** This generator only applies on the nodes for which the decomposition contains part of an X-junction, and expands the decomposition to all four regions forming the X-junction. We build on the assumption that an X-junction results from the boundary of a semi-transparent layer crossing the boundary of two other layers [Sayim and Cavanagh 2011]. This assumption yields a limited set of possible interpretations, which depends on how many of the four regions are already present in the decomposition of the selected nodes:

- None of the regions of the junction have been decomposed. This configuration occurs when we initialize the algorithm, i.e. when the X-junction is used to expand the root of the tree. In the absence of other information, we make the additional assumption that the junction is formed by one semi-transparent layer over two opaque layers, which yields 4 possible configurations illustrated in Figure 5.
- One or more regions of the junction are already in the decomposition. We can deduce that the boundary between an unknown region and a known one is either caused by the end of a known layer or the beginning of a new one. Denoting \( M \) the number of layers in an adjacent known region, we obtain \( M + 1 \) configurations if the unknown region is in the 4-neighborhood of the known one (i.e. the two regions are separated by one boundary), and \( M + 2 \) configurations if the unknown region is in the 8-neighborhood of the known one (i.e. the two regions are separated by two boundaries).

We call the X-junction generator in priority since it allows faster expansion of the tree than the single-region generator.

4.3 Reward

Given the layer assignment of a child node, we can estimate its energy along with the color and opacity gradient function of each layer by minimizing Equation 3. However, this optimization is highly
non-linear, which requires the use of an iterative algorithm like
Newton-Raphson with a good initialization of the solution. We pro-
pose three strategies to avoid computing such a costly optimization
for most visited nodes of the tree.

Layer re-use. Our first strategy to speed up the evaluation of
Equation 3 is to re-use the color and opacity gradients of a parent
node when present in a child node. In other words, we extend the
known gradient functions over the new regions without re-fitting
them. While this approach is approximate, it yields a significant gain
in performance with a marginal loss in visual quality, as evaluated
in Section 5.

Exploitation of X-junctions. As detailed above, expanding the de-
composition at an X-junction only yields configurations with at
most one new semi-transparent layer. In addition, our assumption
on the nature of X-junctions tells us that this semi-transparent layer
covers the boundary between two regions with known layers.
Our strategy is to deduce from this setup a configuration where the
semi-transparent layer is observed over two known colors, which
makes its estimation well-posed [Richardt et al. 2014; Smith and
Blinn 1996].

Denoting \( l \) the unknown layer, each of the two regions on which
it appears gives us equations of the form \( I_l(p) = A_I(p)C_I(p) + (1 - A_I(p))I_{l-1}(p) \) where \( I_l \) is the input image from which we have sub-
tracted all semi-transparent layers above \( l \) and \( I_{l-1} \) is the image
formed by all layers below \( l \), which are known for the two regions
considered. The challenge is now to express these equations over
the same pixels, so that we can remove the term \( A_I(p)C_I(p) \) by
substitution and leave \( A_I(p) \) as the only unknown.

Figure 6 illustrates our algorithm on a toy example, where the
new layer covers two regions \( I^1 \) and \( I^2 \), which share layers with
their known neighboring regions \( I^1_{l-1} \) and \( I^2_{l-1} \) respectively. We
first extend each region over all three other regions so that all
regions share the same spatial domain. Regions that are already
in the decomposition are easy to extend since all their layers are
known and have a parametric form. However, regions covered by
the new layer are still in a bitmap form. Our solution consists in
approximating each such region with a polynomial function using
a least square fit on all pixels. Since each layer is represented by
a linear gradient, a new region can at best be represented by a
polynomial of degree \( M+1 \), where \( M \) is the number of layers shared
with the adjacent regions. We then extend the resulting parametric
functions over the other regions to know their values over all pixels.
We now have the necessary ingredients to form two equations at
each pixel

\[
\begin{align*}
\vec{I}^1_l(p) &= A_I(p)C_I(p) + (1 - A_I(p))\vec{I}^1_{l-1}(p) \\
\vec{I}^2_l(p) &= A_I(p)C_I(p) + (1 - A_I(p))\vec{I}^2_{l-1}(p).
\end{align*}
\]

where \( \vec{I}^1_l \) denotes the extended version of region \( I^1_l \). Subtracting
the two equations gives us an expression where \( A_I(p) \) is the only
unknown

\[
\vec{I}^1_l(p) - \vec{I}^2_l(p) = (1 - A_I(p))\vec{I}^1_{l-1}(p) - \vec{I}^2_{l-1}(p).
\]

Once \( A_I(p) \) is known the minimization of Equation 3 becomes qua-
dratic in \( C_I(p) \) and has a unique solution.

Note that since each region of the image is visited multiple times
during the tree search, we only compute its polynomial representa-
tion at a given degree the first time it is needed, and keep it in
memory for later use.

Sub-sampling. Our last strategy to reduce computational burden is
to use a sparse uniform random sampling of the input image pixels
when fitting the color and opacity gradients of the layers.
However, using very few samples raises the risk of making the
method very sensitive to image noise. Fortunately, we can leverage
the polynomial approximation of the image introduced above as a
means to remove high frequency content that cannot be captured
by the layers. While each polynomial needs to be computed from all
pixels of its region, this is a one time computation which is quickly
amortized over all the nodes where the layers of a given region need
to be estimated.

To sum-up, we first generate for each node a number of config-
urations, which form its child nodes. When a child node does not
contain any new layer, computing its energy \( U \) by Eq. 2 is trivial

Fig. 5. We assume that X-junctions are formed by a semi-transparent layer
crossing two other layers. When none of the layers are known, we assume
that the two bottom layers are opaque, yielding 4 possible configurations.

Fig. 6. Derivation of the opacity of a new layer at an X-junction. Let us
assume that the new layer is shared by the two regions \( I^1_l \) and \( I^2_l \). Fur-
thermore, region \( I^2_l \) shares all of its other layers with region \( I^1_{l-1} \), while region
\( I^2_l \) shares all of its other layers with region \( I^2_{l-1} \). We exploit this redundancy
to deduce the opacity of layer \( l \). We first extend each region over the other
regions to know their values at each pixel (top). Regions \( I^1_l \) and \( I^2_l \) are
trivially extended since their layers are known and have a linear form. We
extend regions \( I^1_l \) and \( I^2_l \) by fitting a polynomial on their pixel colors. We
can then compute a per-pixel opacity \( A \) by simple arithmetic (bottom).
since the parameters of all layers are known. When new layers are involved, the parameters of their color and opacity gradient functions must be estimated first. If the new layer is opaque and below known semi-transparent layers, or if the new layer is part of an X-junction, we can estimate the gradient functions as a quadratic minimization problem that has a unique solution. Otherwise, we run the Newton-Raphson algorithm. Both the quadratic minimization and the Newton-Raphson algorithm are computed with a sparse sampling of the input image.

To further reduce the complexity of the tree, we remove the child nodes whose energies are one order of magnitude higher than the best child node.

4.4 Back-Propagation
Once the energy of new child nodes is computed, we update the tree policy to favor the visit of low energy configurations. We back-propagate information on child node energy towards the root by giving to each node \( t \) the lowest energy of its child nodes

\[
U(x_{t'}) \leftarrow \min_{t_i \in \text{i}^*} U(x_{t_i}).
\]

4.5 Initialization and stopping criterion
We initialize the exploration by generating one root node for each X-junction, using as initial decompositions of these junctions the configurations that minimize \( U_{\text{fidelity}} \). In the cases where no X-junction exists, we ask the user to seed the search by selecting an opaque region in the image. We display a black dot over the selected region of the segmentation for each result where such indication was provided. We stop the exploration once we have reached \( 4N \) leaves of the tree, with \( N \) the number of regions in the image, and keep the decomposition with the lowest energy as our final solution. We adjusted this stopping threshold experimentally by running our algorithm multiple times on the same image using a high threshold, keeping track of the number of leaves visited before reaching the best configuration.

4.6 Finalization
As a reminder, we excluded pixels on the boundary of each region of the segmentation, since those pixels are often corrupted by blur or anti-aliasing. The last step of our algorithm assigns each such pixel to one of its neighboring regions based on goodness of fit. In addition, when an opaque layer is surrounded by another opaque layer, we position the surrounding layer below the surrounded one and extend it to cover the hole, which is occluded by the surrounded layer. As an example, the red layer of the lady bug in Figure 1 extends below the black dots.

Finally, we vectorize the supporting domain of each layer using Potrace [Selinger 2017].

4.7 User interaction
We offer users several means of controlling our method. First, we expose the global \( \lambda \) parameter, which balances fidelity to the input with simplicity of the output. We found that clean bitmap cliparts can be decomposed using a low \( \lambda \) value of 0.1, while photographs often require a higher value of 0.5 to abstract away small details, non-linear color variations, and image noise. We also allow users to increase \( \beta \) to favor the extraction of semi-transparent layers, although we used the default value for all results in this paper.

Second, we allow users to impose that a region of the image be expressed as a single opaque layer. We implement this constraint during the tree search by only generating child nodes with opaque layers over the selected region. We display a black dot over the selected regions of the segmentation for each result where such additional constraint was provided.

Fig. 7. Layered cliparts produced with our method from three studio photographs and a bitmap diagram. We used manual segmentation to properly delineate small and blurry regions (letters on the stop button, highlights on the tomato) and to ignore spurious details (arrows on the chart).
Finally, users can also control the appearance of the decomposition via the input segmentation, for instance to approximate fine details or complex color variations with a unique region.

5 EXPERIMENTS

Figure 1 and Figure 7 show results of our method on a variety of illustrations and studio photographs. Our approach is especially effective at expressing highlights and shadows with semi-transparent layers, which facilitates subsequent manipulation like color and shape editing, as shown in Figure 1(c).

We now compare our method to existing work in the field and evaluate the impact of parameters and performance.

5.1 Comparisons with existing work

Figure 9 compares our method with the work by Tan et al. [2016] and Richardt et al. [2014]. The method by Tan and colleagues targets the different application of bitmap color editing. As a result, while their method succeeds in separating the image into layers based on the dominant colors of the image, each layer is a bitmap with a constant color and spatially-varying transparency rather than a vector path filled with a parametric gradient.

Our goal and approach is more similar to the work by Richardt et al., and our method produces similar results to theirs. The main difference between the two methods resides in the user workflow. The interactive workflow of Richardt et al. requires users to extract each layer of the decomposition one by one, in a front to back order. Since their layer extraction takes between a few seconds to a minute per layer, their results took between several minutes to an hour to create, as detailed in Section 5 of their paper. In contrast, users of our system simply have to provide a segmentation of the image, which takes a few minutes with an interactive tool, and optional indications of opaque regions before letting our algorithm produce the entire decomposition within seconds. Another difference resides in the family of gradients supported by the two methods. Our method extracts 2-stop gradients, while Richard et al. also support 3-stop gradients, but requires users to indicate the number of stops for each region. This is why their method extracts each color band of the cone as a single 3-stop gradient while we extract them as two 2-stop gradients (Figure 9, 3rd row). In addition, Richard et al. represent opaque layers with gradient meshes, while we only use linear gradients. Their gradient mesh better captures contrast on the opaque layer of the cylindrical mug (Figure 9, 4th row).

5.2 Impact of parameters

The main parameter of our algorithm is \( \lambda \), which controls the amount of abstraction of our vectorization. Figure 8 details its impact on the vectorization of a bitmap clipart. A low value of \( \lambda \) reproduces the input as closely as possible with linear gradients. Increasing \( \lambda \) favors the use of semi-transparent layers at the price of more approximate rendition of the original colors. A high value of \( \lambda \) abstracts away details in an effort to reduce the number of layers.

5.3 Impact of pre-segmentation

Figure 10 compares the output of our method with different input segmentations generated with a Mean Shift algorithm [Comaniciu and Meer 2002] and a manual segmentation. Our method produces visually similar vectorizations with two different automatic segmentations. In particular, multiple regions are merged to form a few layers thanks to our simplicity term. However, we obtain more stylized cliparts using manual segmentations.

5.4 Performance

Table 1 provides timings for several of our results, measured on a computer equipped with a 3rd generation core i7 processor and 16Gb DDR3 memory. Our algorithm took less than a minute to produce each decomposition, which allows an interactive workflow where users can refine the result by adding a new opacity constraint or modifying the segmentation if necessary. The last column of Table 1 detail the number of opaque regions selected by the user to initialize the optimization.
Fig. 9. Comparison with existing decomposition methods. The layers generated by [Tan et al. 2016] are bitmaps filled with a constant color and spatially varying transparency, which is suitable for color editing but not for other applications of vector graphics. Our results are similar to the ones by [Richardt et al. 2014], although our method requires less user intervention and solves for the layer parameters more efficiently. All our results were produced with a manual segmentation, except the cone (3rd row).
Table 2 compares several versions of our algorithm where we removed important features. In particular, exploiting X-junctions greatly speeds up the exploration of the solution space, while increasing the maximum number of child nodes does not yield a significant increase in quality. We also implemented re-fitting of all layers for each child node, which increases timing by two orders of magnitude without much impact on the energy.

### 5.5 Limitations

Our choice of restricting the image representation to linear gradients is key to achieve joint abstraction and vectorization. However, our current implementation does not support variants of linear gradients, such as radial gradients, although such primitives could be included in the configurations evaluated by the tree search. More advanced primitives like gradient meshes would be more difficult to integrate since they cannot be trivially expanded to adjacent regions, which is a key assumption made by our algorithm.

Our method targets piecewise-smooth images and as such reaches its limits in the presence of texture. A low-contrast texture is averaged-out if segmented as a single region, while high-contrast textures can be segmented in multiple small regions, which result in a cluttered clipart with many layers. The house in Figure 10 illustrates how our approach performs on a complex natural image with textures.

We speed up our algorithm by assuming that X-junctions are caused by transparency. This assumption breaks on texture patterns such as a checker-board, where many X-junctions are not
While image vectorization has received significant attention in the computer graphics community, very little work has been done on reproducing the style and layer structure of traditional vector cliparts. Motivated by the ubiquity of linear gradients in vector art, we have presented an algorithm based on Monte Carlo Tree Search to jointly decompose an image into layers and approximate these layers with linear gradients.

Our method takes a segmented image as input and does not attempt to modify the shape of the segmented regions, apart from fusing small segments to crate bigger ones. However, the regions produced by automatic segmentation algorithms often have more intricate shapes than the ones created by vector artists. An interesting direction for future research would be to jointly simplify shape and color during the vectorization process, potentially by including a shape simplicity term in our energy formulation.

**REFERENCES**


