Music2Image: A synesthetic approach for music perception

ABSTRACT

We present a synesthetic approach that uses visual effects created on a given image to improve the music comprehension. Synesthesia means a simultaneous perception of different sensorial stimulus.

In this direction, our approach combines visual stimulus of video with sonorous stimulus of music. The synesthetic approach is performed by the appropriate use of image operators (image processing techniques) to create different effects in an image. These operators are used to highlight the occurrence of some events in the music.

The main goal of this work is to improve perception and cognition of music elements. Chosen images and operators are a combination of aesthetic values and musical elements that shall be highlighted.

All videos related to the following experiments are available at: https://eden.dei.uc.pt/∼vitorgr/Music2Image.html

1. INTRODUCTION

In this work, an approach for creating visual effects in a given image which stimulates the perception of music elements (like timbres, chords, notes and velocity) is introduced. Simultaneous perception of different stimulus reinforces the occurrence of such events. The approach presented in this paper is based on this principle.

The main goal of our method is to increase the comprehension of music elements (a sonorous sensory stimulus) through visual elements in an image (a visual sensory stimulus). Thus, a synesthetic approach which combines seamlessly music and visual effects (along a video) is introduced. The visual effects are performed by the use of image operators, i.e., image processing techniques used to highlight the occurrence of events in the music while it is being played.

We believe that aesthetic values are important in this context. Thus, we will present examples which satisfy it. However, the focus is on approaches in which the methodologies enhance the cognitive aspects between music and image. Thus, despite of this work can be used for artistic purpose, our main goal is related to the perception and cognition of the music elements.

The relationship between music and image can be established by the association of notes played in a musical piece

and the color of pixels in the image. Those colors can be the original or defined with a rule for creating a visual effect that increases the comprehension of a music element. Such relationship is created by the adequate use of image processing operators that must be controlled through events related to the music.

We are presenting a synesthetic approach that uses visual effects performed on an image to highlight events of a song. The use of images as the basis for visual effects adds an aesthetic value to music representation (since the image has intrinsically this aesthetic value) without losing cognitive value. However, it is necessary to guarantee that the visual effect satisfies certain principles. Such principles are discussed in this paper.

In Section 2, we will present some related works, and we will highlight the synesthetic features of them. In Section 3, we will introduce our approach. First, we will talk about the relation of time in music and in video. After, we will explain how to relate the elements in music with events on video. Then, we will characterize the image operators, and explain how to use them. In Section 4, we will present some of our experiments, talking about the creation of operators and their synesthetic purpose. Finally, in Section 5, we will discuss about our results and conclusions.

2. RELATED WORKS

Galeyev [1] defines synesthesia as intersensory association formed by similarity of heteromodal perceptions, the result of a simultaneous associative process. The author argue that synesthesia is an essential component of musical thinking.

Higara et al. [2] propose a visualization system to create synesthesia between visual and audio information. The authors describe which information should be shown by the system to maximize expressive performance of different instruments. While Higara's work aims to visually express the different instruments in a piece of music, our work aims to express other music elements, such as: timbre, pitch, and volume using image operators.

Lewiner et al. [3] propose music visualization through 3D mesh deformation. The control is based on parameters extracted from music spectral analysis. In our work, the image operators are related to parameters extracted from a Musical Instrument Digital Interface (MIDI) file. Given that, we can develop more complex images operators because we know more information about the musical composition through the MIDI file.
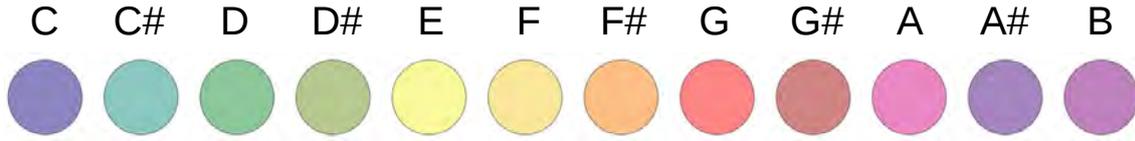
Figure 1. Sound tones to colour tones according to Louis Bertrand Castel [4].

3. THE METHOD

To create a video for a music, we define an image operator parameterized for certain elements of the music. For each frame generated in the video, it is necessary to identify the notes that are being played for each instrument (timbre), as well as, their velocity (volume), and then, to generate the respective representative visualization.

In this work, we will not discuss song segmentation. Thus, if a segmentation is provided, it is easy to associate an operator for each segment in the song, which increases the dynamism of the presentation. Another possibility is to highlight different elements in each segment (and also, to highlight the segmentation itself). Besides that, even within a song segment it is possible to change operators, or change the image that is being utilized. Another possibility is to synthesize an image from scratch, given the music events (we will not discuss such approach in this work, but it is perfectly compatible with our approach).

To create a video containing a visual effect related to a given music, it is necessary to establish a relation between the time in music (tick) and the time in video (frame). This topic will be presented in Subsection 3.1. Furthermore, in Subsection 3.2, we associate events of music, related to timbres, chords and notes, with events of video. Finally, in Subsection 3.3, we will distinguish the image processing operators we can use for the purpose of this work.

3.1 Time on Music and on Video

In our experiments, the music is described in a MIDI file [5] that contains which notes will be played, defining, for each note, the respective instrument (timbre), the tick on (t_i) and the tick off (t_f), as well, the note type and the velocity the note will be played. It is important to highlight the music's unit time is tick.

On the other hand, the creation of a video consists in defining several frames that will be exhibited according to a pre-defined frame rate. Thus, for creating the k-th frame, we have to define the respective time on music (in tick) that is happening at the same time.

We have to pre-set the rate of Ticks Per Second (TPS), which depends on the time signature of the music. Thus, to know the duration in seconds of music with N notes, we have to multiply the tick off of the last note t_f by the TPS constant, i.e, $T = t_f * TPS$. Furthermore, since the FPS constant was set, we can define how many frames it is necessary to create the video, by $F = T * FPS$. Analogously, we can define the tick of the j-th frame by $tick(j) = \frac{j}{TPS*FPS}$. So, we can decide which notes are being played in this frame by checking which notes satisfying $t_i \leq tick(j) \leq t_f$.

3.2 Music elements and Video events

In this section, we will present our approach for creating the relation of the music's elements with effects in the frames of a video. In particular, we will create a video style by determining for each frame what is the color of each pixel according to the music's elements (the timbre, the chords, the notes and the velocity of the note is being played) are being played in the respective time. The association of music with video is performed by triggering events when a new frame is created and when a new chord is played.

The relation between video events with music elements is established by the adequate use of image operators, i.e., the visual effects created for a video is based on the association of image operators controlled according to information associated to the music. In the next section, we will present how these image operators are and how to use them.

3.3 Image Operators

We call by image operator a function which input and output are an image (in our case, both images have the same resolution). Some examples of image operators are classical image processing techniques, like: filtering, clustering, warping, morphing and image combination. Furthermore, we can define some specific image operators, for instance, by creating rules for changing the color (luminance and chrominance) of some image's pixel according to some specific music event. Indeed, we can define more complex image operators by combining some of these techniques.

Hence, we will present some examples of video style defined by some specific image operators, each one associated with some music elements in a particular way. Because of this, it is the most flexible part of this work, and we can use many operators for producing different effects for highlighting different music elements.

Not all image operators will produce a synesthetic effect. Nevertheless, we can define some principles to guide their creation:

- Associate between colors and notes

- Relate the amount of processed pixels and the velocity of the note

- Split the frame area in regions for each timbre

In next section, we will describe some experiments created by following those principles. For creating these video styles, we used the color palette of Louis Bertrand Castel [6] to establish the relation between the color and the notes. This palette is shown in Figure 1. Despite of this choice being arbitrary (i.e., we could choose any other palette color), it makes sense because its synesthetic effect was tested in some other researches [7].

Figure 2. Luminance experiment: we change the chrominance of pixels inside of circles, related to the notes being played, using the note color and keeping the luminance.

4. EXPERIMENTS

In the following experiments, we use two songs: the Bach's Minuet and the Ravel's Boléro. The first song is played by a piano (only one timbre), and the second one is played for a string quartet (four timbres). Both songs are played in the 4x4 time signature, and thus, we set the TPS as $\frac{0.5}{256}$. And, both MIDI files were extracted from the *www.8notes.com*.

Furthermore, the FPS was set as 24, for all videos. All images used are under the creative commons license, and were extracted from Flickr dataset.

All videos related to the following experiments are available at: *https://eden.dei.uc.pt/~vitorgr/Music2Image.html*.

4.1 Luminance

In this experiment, the visual identity changes every time a new chord is played. We only show information related to the chords which are being played.

For creating this visual effect, we start with a grey scale image and for each note being played we highlight a region inside of a circle. The color into the circle depends on the type of the respective note. Similarly, the size of this circle depends on the velocity of this note. When a chord has more than one note, we show a circle for each note. Figure 2 ilustrates this visual effect.

For each timbre, we define a respective region into the image. The position of circle related to a note is randomly chosen into the region associated with the respective timbre. For example, when the music is played using only one instrument, the timbre region is the entire image. In the case of music played with four instruments, we can divide the image in four congruent rectangles, and set each one for a respective timbre. This division approach was used in all experiments.

Another possibility to highlight the difference between timbres is to use different shapes for each one. In Figure 4, we show an example of a song played for four instruments, and we use different types of polygons for each one.

Since the position and size of circles are chosen, we change the chrominance of the pixels into these circles. We set the chrominance with the respective color in the note palette and keep the luminance. This example is very simple but illustrates the use of image operator principles. In next subsections, we will present more sophisticated methods.

4.2 Image Chrominance

This experiment is an extension of the previous one. In this case, we also start the visual effect from a grey scale image, and every time a note is played we show a circle with the respective chrominance of the note. Nevertheless, for this visual identity we created an effect continuous on time. As far as time goes, the color of the pixels inside of the circles varies from the note color to grey (with the respective pixel luminance) passing intermediately by the original color of the image.

For such purpose, we create for each new chord an image, denoted by N, similarly to the one created in previous experiment (changing the chrominance into the circles). Furthermore, we create another image L, that it is initialized on the new chord event by the previous frame and updated as $L = (1 - \lambda)L + \lambda I$, where λ is a parameter stored in a grid with the same resolution of frames. We call this grid by Lambda Map. In the lambda map, we increase all points into the circles of new notes for a constant δ. We set this $\delta = 150$. For each new frame, we decrease a small value on all points of lambda map (we set this another value by 4).

After that, for each frame, lets t_i and t_f the initial and final ticks of the notes are being played, respectively, we calculate $t = \frac{tick - t_i}{t_f - t_i}$. With this parameter, we create the frame interpolating L and the original image using t: $F = (1 - t)L + tI$.

This interpolation guarantees the desired effect, because when $t = 0$, we have the color notes, and when $t = 1$ we

(a) Frame
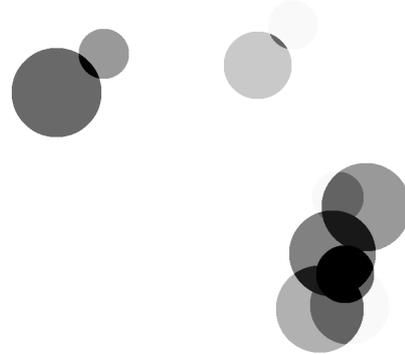
(b) Lambda Map

Figure 3. Image Chrominance experiment: we interpolate the colors of the note circles starting from the note colors, passing to the image color (using the lambda map), and returning to respective grey (luminance).

Figure 4. Image Chrominance experiment using four timbres.

have the image colors. After the time interval, when the note is played, the lambda value will decrease until zero, and thus the color will return to grey (this effect will happen in image L during the following notes). The visual effect of this operator is shown in Figure 3.

An advantage of this approach is to have a memory of the notes recently played. It happens because there are points outside of the circles of notes being played which the lambda value is greater than zero (and thus the color will be shown). It creates a semi-global effect (i.e., we can see information of what is happening in the respective time, but also, what happens a bit before). Furthermore, the decreasing of the lambda map values makes this memory limited, and hence, there is not too much information that can confuse the observer.

4.3 Show Pixels

Given a music, we define which types of notes will be played (valid notes). Next, for each pixel we choose the note which color (according to Castel's palette) is more similar to one of valid notes, and then, associate this pixel to this note. This process creates an image clustering according to the color of valid notes.

While music is being played, every time a new note is played we show a fraction of the total number of pixels associated to this note. This fraction is the rate of the respective amount of pixels and the amount of times it shall be played.

To guarantee the amount of pixels shown for all note types are similar, it is necessary that the previously mentioned fraction is almost the same for all valid notes. To achieve this goal, it is necessary to search an image satisfying this property (unless a threshold). In our experiments, we search for this image in a dataset with 60 thousand of exemplars, and we found tens of good images for all songs.

Another way to associate pixels and notes is to weight the incidence of a note by its velocities. In this case, the rate of note is the respective amount of pixels divided by the sum of the velocity of all times it is played. In this way, instead of to show a steady amount of pixels per note, we can multiply the velocity by the rate of this note. This approach satisfies the image operator principle.

Furthermore, we add an interpolation of colors during the

time interval while the note is shown. The color begins with the note color and ends with the image color. This interpolation is performed using the parameter t, similarly what was explained in previous subsection.

The choice of which pixels will be shown is a sensitive point for obtaining an adequate synesthetic effect. For instance, if the pixels are randomly chosen (between those possible) they are spread along the image. Because of this scattering, it is not easy to understand the meaning of the visual effect. However, this problem is easily solved by agglutinating the shown pixels in the image. This solution creates the desired synesthetic effect. Figure 5 shows some frames created by this operator.

4.4 Image Combination

In this experiment, we combine two images to create this visual effect. The first is shown in foreground and it is not related to the music (according to the pixels and notes criterion previously described). The second image is in background, hidden by the first one, and it is desired to be related to the music. In this way, when the music is played, we open windows in the first image to show parts of the other image related to notes being played.

This window is a circle, which size depends on the velocity of the note. Inside of this window, pixels related to the note are exhibited with its original color. The others are shown with its luminance and the chrominance of note (similarly, the experiment of luminance).

This method has an easy puzzle for discovering the hidden image. We believe it is a stimulus for the observer, which helps to attract attention for the presentation. It will not distract him of the main goal (understand the music) because we use large windows that helps to understand easily the hidden image.

This operator is less sensitive to the image choice than the previous one. Here, if the chosen image is not related to the music, what is exhibited inside of the note circles is almost the luminance and note chrominance representation. On the other hand, if the image is well related to the music, this area will be almost equals to the original image.
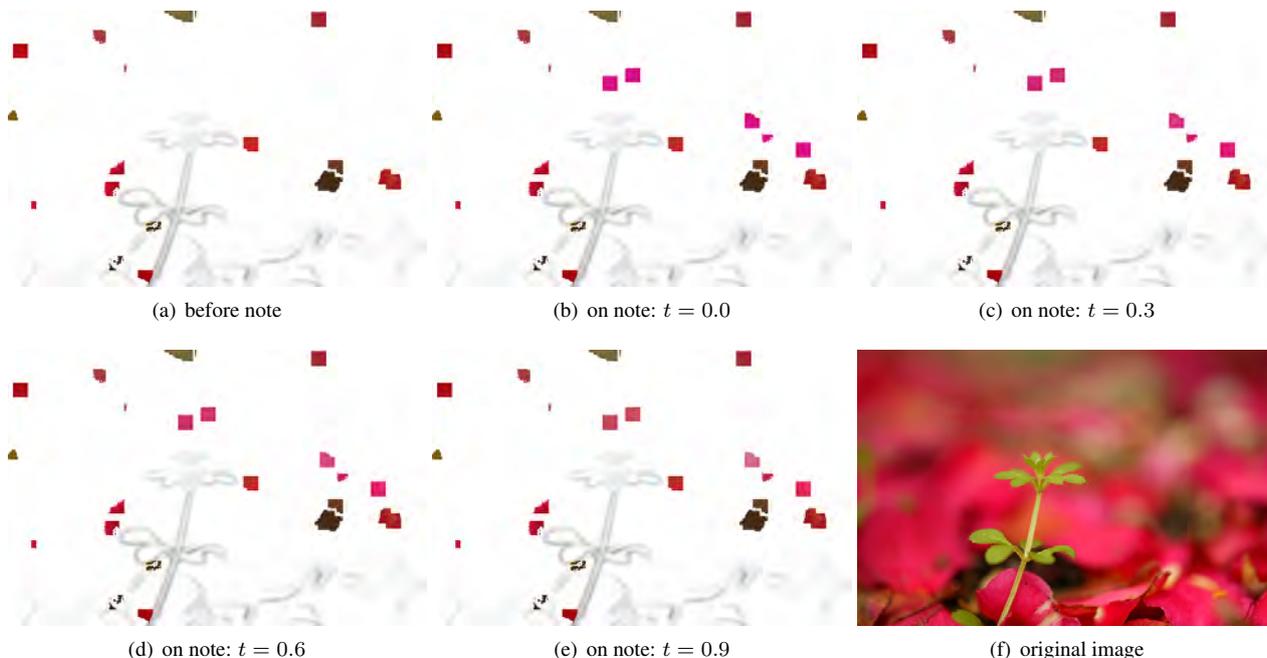
(a) before note

(b) on note: $t = 0.0$

(c) on note: $t = 0.3$

(d) on note: $t = 0.6$

(e) on note: $t = 0.9$

(f) original image

Figure 5. Show Pixels experiment: we show some pixels related to the notes are being played.

5. RESULTS AND CONCLUSIONS

We present three principles used to relate an image with elements of music which can be used to define some image operators. These operators create visual effects in the image that improve the comprehension of music. This approach has a synesthetic characteristic that is useful to increase the perception and the cognition of music elements. Therefore, it can be used to improve the music appreciation and to help the process of musical learning.

Our approach for visual effects creation is based on images. Our images can be photographies, textures, or synthetic images. Another possibility that fits in our approach is to use the music to control an image synthesis method. We did not test this approach, but we believe it is a good direction for future work.

Some image operators, introduced in this work, depend on the choice of an adequate image. This choice aggregates an aesthetic value to the presentation, and thus, the use of these methods sensitive to the matching between image and music depends on the availability of a large image dataset. Fortunately, this is not a problem nowadays, since there are datasets like Flickr, providing many images under creative commons license that can be massively downloaded. Therefore, the problem is constrained to choose a subset of images, in this large dataset, according to a given music, and then to pick manually one of them with a good aesthetic (or with some other special feature).

Finally, besides of the mentioned operators, we can create many others by combining different image processing techniques to highlight specific features on the music. A future work is to create other image operators, and possibly to extend the principles list (to add other music elements). Moreover, we intend to improve the evaluation of the synesthesia in our method, through a survey with users.

6. REFERENCES

[1] B. M. Galeyev, "The Nature and Functions of Synesthesia in Music," *Leonardo*, vol. 40, no. 3, pp. 285–288, May 2007.

[2] R. Hiraga, R. Mizaki, and I. Fujishiro, "Performance Visualization: A New Challenge to Music Through Visualization," in *Proceedings of the Tenth ACM International Conference on Multimedia*, ser. MULTIMEDIA '02.   New York, NY, USA: ACM, 2002, pp. 239–242.

[3] T. Lewiner, C. Marques, J. Paixão, S. Botton, A. Cabral, R. Nascimento, V. Mello, A. Peixoto, D. Martinez, and T. Vieira, "Stereo music visualization through manifold harmonics," *The Visual Computer*, vol. 27, no. 10, pp. 905–916, 2011.

[4] K. Brougher, J. Strick, A. Wiseman, and J. Zilczer, *Visual Music: Synaesthesia in Art and Music Since 1900*. Thames and Hudson, 2005.

[5] G. Loy, "Musicians Make a Standard: The MIDI Phenomenon," *Comput. Music J.*, vol. 9, no. 4, pp. 8–26, dec 1985.

[6] T. Hankins and R. Silverman, *Instruments and the Imagination*.   Princeton University Press, 1995. [Online]. Available: http://www.jstor.org/stable/j.ctt7ztvhb

[7] J. H. Fonteles, M. A. F. Rodrigues, and V. E. D. Basso, "Creating and Evaluating a Particle System for Music Visualization," *J. Vis. Lang. Comput.*, vol. 24, no. 6, pp. 472–482, Dec. 2013.
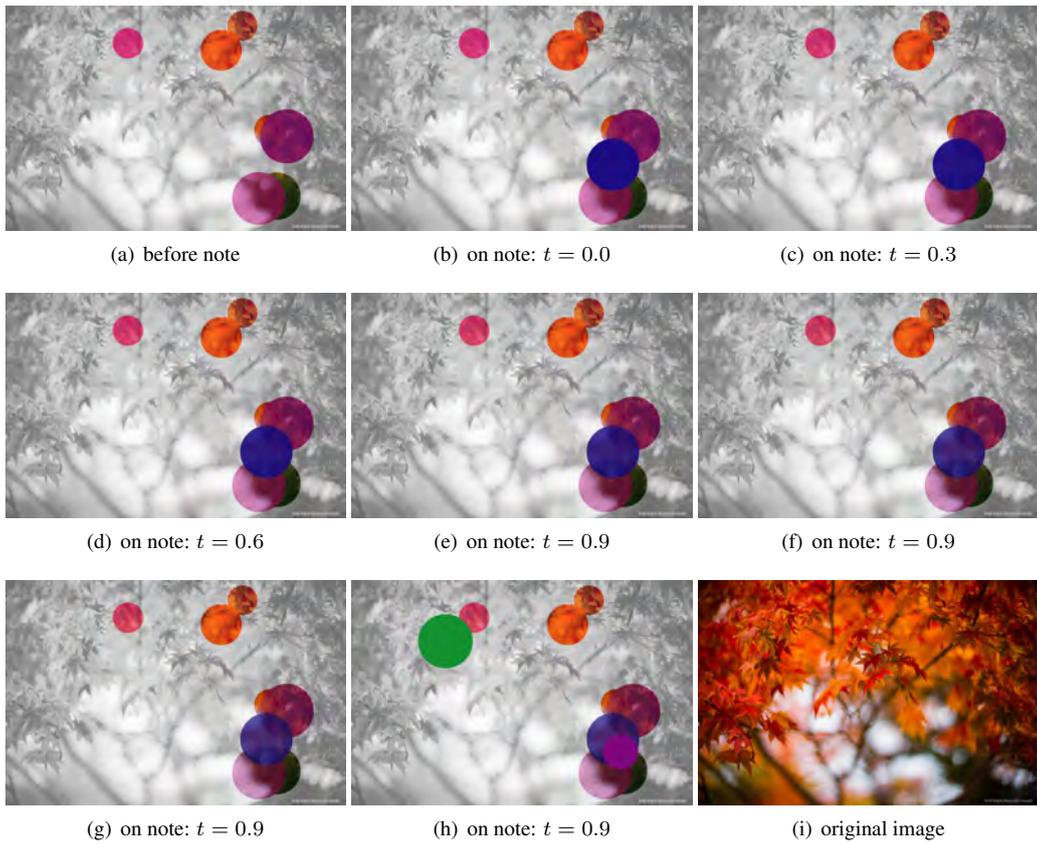
(a) before note   (b) on note: $t = 0.0$ (c) on note: $t = 0.3$

(d) on note: $t = 0.6$ (e) on note: $t = 0.9$ (f) on note: $t = 0.9$

(g) on note: $t = 0.9$ (h) on note: $t = 0.9$ (i) original image

Figure 6. Some frames of the Image Chrominance experiment.

(a) original image   (b) luminance image   (c) foregrounf image
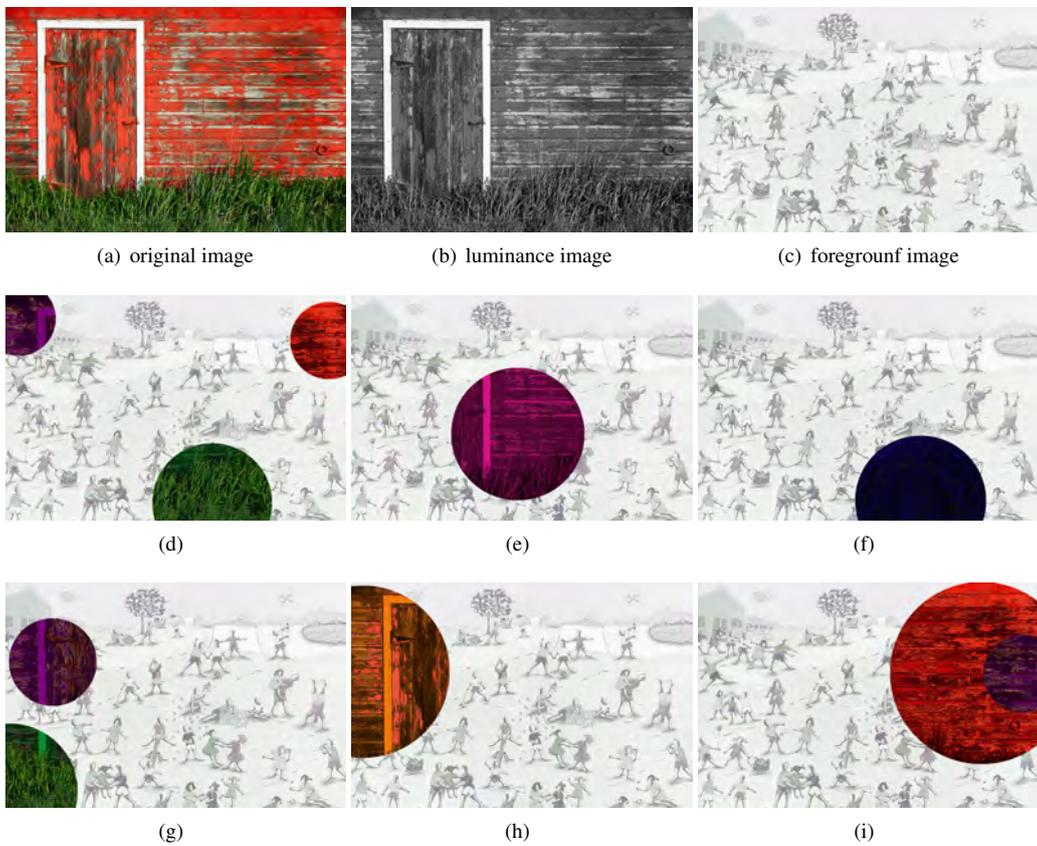
(d)   (e)   (f)

(g)   (h)   (i)

Figure 7. Some frames of the Image Combination experiment.