Motion-Compensated Frame Interpolation on Contrast Enhanced Video

Jean P Johny
Federal Institute of Science and Technology (FISAT)
Mookkannoor

Arunkumar M N, Ph.D.
Assistant Professor, Computer Science Department, FISAT

Sira Salim
Federal Institute of Science and Technology (FISAT)
Mookkannoor, Angamaly

Roopa Gokul
Federal Institute of Science and Technology (FISAT)
Mookkannoor, Angamaly

ABSTRACT
Contrast enhancement plays an important role in the improvement of visual quality for computer vision, pattern recognition, and the processing of digital images and videos. Efficient contrast enhancement using adaptive Gamma correction with weighting distribution is an efficient method to modify histograms and enhance contrast in videos. A major problem faced by this approach is the discontinuity in the resulting enhanced video due to some missing frames. To cope up with this problem we merge the concept of interpolation with contrast enhancement. Here we use a motion-compensated frame interpolation (MCFI) algorithm to increase video temporal resolutions based on multihypothesis motion estimation and texture optimization. Thus the proposed method improves the brightness of dim videos via the gamma correction and probability distribution of luminance and MCFI algorithm preserves the continuity of the enhanced video.

General Terms
Image Processing, Contrast Enhancement

Keywords
Gamma correction, Contrast enhancement, Histogram equalization, Motion compensated frame interpolation.

1. INTRODUCTION
Video enhancement process consists of a collection of techniques that seek to improve the visual appearance of the video or to convert the video to a form better suited for analysis by a human or machine. Contrast Enhancement is essential to improve the quality of the videos that are captured in extreme lighting conditions, such as excessively bright or dark environments that produce low contrast, which produce normal global contrast videos with a low dynamic range in shadowed areas

Contrast enhancement plays a crucial role in image processing applications, such as digital photography, medical image analysis, remote sensing, LCD display processing, and scientific visualization. There are several reasons for an video to have poor contrast: the poor quality of the used device, lack of expertise of the operator, and the adverse external conditions at the time of acquisition. These effects result in under-utilization of the offered dynamic range. As a result, such videos may not reveal all the details in the captured scene, and may have a washed-out and unnatural look. Contrast enhancement targets to eliminate these problems, and to obtain a more visually pleasing or informative video. The visual appearance of video may be significantly improved by emphasizing its high frequency contents to enhance the edge and detail information in it. The principal objective of video enhancement is to modify the attributes of the video to make it more suitable for a given task and a specific observer. During this process, one or more attributes of the video are modified.

One of the most popular global contrast enhancement techniques is histogram equalization (HE) [3]. It flattens and stretches the dynamic range of the frames histogram and results in overall contrast improvement. Efficient contrast enhancement using adaptive Gamma correction with weighting distribution is an efficient method to modify the histograms and enhance contrast in videos. Gamma correction techniques make up a family of general HM techniques obtained simply by using a varying adaptive parameter The simple form of the transform-based gamma correction (TGC) is derived by

$$T(l) = l_{max}\,({}^{l}/_{lmax})^{\gamma} \qquad (1)$$

where $l$max is the maximum intensity of the input. The intensity l of each pixel in the input frame is transformed as T(l) after performing this.

However, when the contrast is directly modified by gamma correction, different frames will exhibit the same changes in intensity as a result of the fixed parameter. Fortunately, the probability density of each intensity level in a frame can be calculated to solve this problem. The probability density function (pdf) can be approximated by

$$pdf(l) = n_l/(MN) \qquad (2)$$

where n_1 is the number of pixels that have intensity l and MN is the total number of pixels in a frame.

The cumulative distribution function (cdf) is based on pdf , and is formulated as:

$$cdf(l) = \sum_{k=0}^{l} pdf(k) \qquad (3)$$

After the cdf of the frame is obtained from equation (3), traditional Histogram Equalization (THE) directly uses cdf as a transformation curve expressed by

$$T(l) = cdf(l)l_{max} \qquad (4)$$

But the main problem with video contrast enhancement is that the resulting video is not complete. Many of the frames in the video will be missing and thereby it is not continuous. A video sequence with a low frame rate incurs motion aliasing and blurring artifacts. It is uncomfortable to watch such a sequence. Thus, it is essential to replace the missing frames and to develop efficient frame rate up to increase temporal resolutions and provide smoother transitions between frames. So in order to overcome the discontinuity problem and to make smoother videos we use interpolation algorithm along with the contrast enhancement technique. Recent frame rate up techniques exploit the motion information in video sequences to achieve faithful interpolation results. Here a motion-compensated frame interpolation (MCFI) algorithm [2],[7][10] is used to increase video temporal resolutions based on multihypothesis motion estimation and texture optimization. A typical MCFI scheme consists of two processes: motion estimation and frame interpolation. In other words, MCFI first analyses each pair of consecutive frames to estimate motion vectors (MVs) and then interpolates the intermediate frame by halving those MVs.

The rest of the paper is ordered as follows. In section 2 we discuss some related works. Section 3 describes the proposed framework in detail. Section 4 presents the conclusion.

2. LITERATURE SURVEY

Several methods for contrast enhancement have been proposed in the recent years. Arici. T et.al proposed a histogram modification framework and its application for image contrast enhancement in 2009[3]. The presented framework employs carefully designed penalty terms to adjust the various aspects of contrast enhancement. Hence, the contrast of the image/ video can be improved without introducing visual artifacts that decrease the visual quality of an image and cause it to have an unnatural look. Eventhough it does not produce any artifacts as Histogram Equalization and weighted threshold HE its time complexity is worse. Kim. M et.al proposed a new histogram equalization method, called RSWHE (Recursively Separated and Weighted Histogram Equalization), for brightness preservation and image contrast enhancement [4]. The essential idea of RSWHE is to segment an input histogram into two or more sub histograms recursively, to modify the sub-histograms by means of a weighting process based on a normalized power law function, and to perform histogram equalization on the weighted sub-histograms independently. This approach produces better image quality and displays more information the image holds but it shows some unnatural high intensity in some images. Later Lee. C et.al proposed A powerconstrained contrast-enhancement algorithm for emissive displays based on histogram equalization (HE) [6]. They first propose a log-based histogram modification scheme to reduce overstretching artifacts of the conventional HE technique. Then, they develop a power-consumption model for emissive displays and formulate an objective function that consists of the histogram-equalizing term and the power term. By minimizing the objective function based on the convex optimization theory, the proposed algorithm achieves contrast enhancement and power saving simultaneously. But this approach decreases the overall brightness of the input image. Also it reduces the contrast for infrequent input-pixel values and for some images it sacrifices the details.

Recently several Frame rate up-conversion (FRUC) techniques have been proposed by various researchers. FRUC techniques exploit the motion information in video sequences to achieve faithful interpolation results. Thus, they are also referred to as motion-compensated frame interpolation (MCFI) [8]. MCFI first analyses each pair of consecutive frames to estimate motion vectors (MVs) and then interpolates the intermediate frame by halving those MVs. Unnatural shapes and textures on the interpolated frame was corrected by Exemplar-based texture synthesis techniques which corrects erratic textures after MCFI by employing the texture optimization technique [9]. Although the exemplar-based FRUC algorithm preserves static structures faithfully in an interpolated frame, it often falls into a local minimum and distorts dynamic objects. This algorithm cannot be applied for natural image synthesis, since many different textures coexist within a natural image. Hence, it is necessary to constrain the source-target matching to be performed within the regions of the same texture. The block matching algorithm often fails to find the robust correspondence between blocks including multiple objects with disparate motions. Such failures cause significant blurring artifacts in interpolation results. This problem can be overcome by employing the graph-based video segmentation technique in [10] to measure the reliability levels of MVs and improve their accuracies.

In this work, we propose a novel framework for video enhancement. The proposed framework employs video contrast enhancement based on gamma correction with weighting distribution and subsequent motion-compensated frame interpolation based on multihypothesis motion estimation for avoiding the discontinuity of resulting video.

3. PROPOSED METHODOLOGY

To enhance the video sequence, our proposed method uses adaptive gamma correction. First, the video sequences are converted into frames. These frames are processed using AGCWD method [1]. To remove the discontinuity among the frames, a novel motion-compensated frame interpolation (MCFI) algorithm [2] is applied.

3.1 Proposed Algorithm

1. The original input video is converted into frames.
2. First incoming frame is directly stored in the frame storage.
3. Then this frame is used to generate a mapping curve for the AGCWD method.
4. For subsequent incoming video frames, entropy model is used to measure the differences of the information content between two successive frames.
5. When the absolute difference between the current H and previous H exceeds threshold Th, the frame storage can be updated by the incoming frame, while the transformation curve is also modified.
6. Modified frame is used to generate enhanced video.
7. To remove the discontinuity among the frames, multihypothesis motion estimation interpolation is applied over the video.

3.2 AGCWD Method

To enhance the video sequence, our proposed method uses adaptive gamma correction. In our method a hybrid HM method is used, which combines the Traditional Gamma Correction and THE (Traditional Histogram Equalisation) methods. The adaptive gamma correction (AGC) is formulated as follows:

$$T(l) = l_{max}\,(\frac{l}{l_{max}})^{\gamma} = l_{max}\,(\frac{l}{l_{max}})^{1-cdf\,(l)} \qquad (5)$$

The AGC method can progressively increase the low intensity and avoid the significant decrement of the high intensity. Furthermore, the weighting distribution (WD) function is also applied to slightly modify the statistical histogram and lessen the generation of adverse effects. The WD function is formulated as:

$$pdf_w(l) = pdf_{max}\,(\frac{pdf\,(l)-pdf_{\,min}}{pdf_{max}-pdf_{\,min}})^{\alpha} \qquad (6)$$

where α is the adjusted parameter, $pdf_{\max}$ is the maximum pdf of the statistical histogram, and $pdf_{\min}$ is the minimum pdf. The modified cdf is:

$$cdf_w(l) = \sum_{l=0}^{l_{max}} pdf_w(l) / \sum pdf_w \quad (7)$$

where the sum of $pdfw$ is calculated as follows:

$$\sum pdf_w = \sum_{l=0}^{l_{max}} pdf_w(l) \quad (8)$$

Finally, the gamma parameter based on *cdf* of Equation () is modified as follows:

$$\gamma = 1 - cdf_w(l) \quad (9)$$

The method is applied to the video frames in HSV color model. In the HSV color model, the hue (H) and the saturation (S) can be used to represent the color content, with the value (V) representing the luminance intensity. The color video frame can be enhanced by preserving H and S while enhancing only V. AGCWD method was applied to the V component for color contrast enhancement. Fig 1 shows the overall AGCWD method.

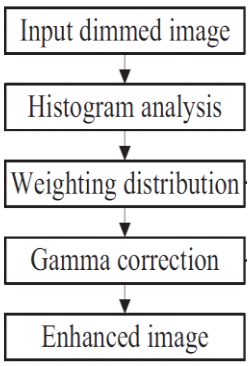


Fig 1: Flowchart of the AGCWD method.

3.3 Motion-Compensated Frame Interpolation (MCFI)

The algorithm is based on the MHME (multihypothesis motion estimation) and the texture optimization. The algorithm makes multiple motion hypotheses for each pixel using different ME parameters. Then, it solves the labeling problem to optimize the parameters and determine the best motion hypothesis for each pixel. More specifically, the labeling cost takes into account the color and segmentation information in the source frames to evaluate the reliability levels of the motion hypothesis. By minimizing the labeling cost, the algorithm obtains a reliable motion hypothesis field. It refines the motion hypothesis field based on the texture optimization technique and blends multiple source pixels to reconstruct each pixel in the intermediate frame.

3.4 Multi Hypothesis Motion Estimation (MHME)

First, for each pixel, we construct multiple motion hypotheses by estimating multiple MVs with different parameters. Each ME parameter is a vector representing the block size and the direction. Second, video segmentation results are used to detect unreliable motion hypotheses and refine them. Third, formulate the parameter optimization as a labeling problem and employ the global energy minimization technique to obtain the best motion hypothesis for each pixel.Fig.2 shows an overview of the MHME algorithm.

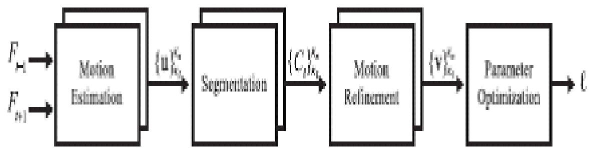


Fig 2: An overview of MHME algorithm

1. Motion Hypothesis: Let k Є K be the parameter representing the block size and the direction of the ME, and let $B^{(k)}_p$ denote the block containing p when the parameter is k and we use the forward or backward ME. The MV $u^{(k)}_p$ for pixel p with parameter k is estimated by minimizing the matching cost J. In other words, $u^{(k)}_p = argmin_u$ J (u; $B^{(k)}_p$) For each parameter k Є K, we estimate the MV $u^{(k)}_p$ and the corresponding MV v^k_p for p'in the intermediate frame F_t. Some pixels in F_t may not be assigned MVs, yielding holes. In such a case, as the MV of each hole pixel, we employ the average MV of neighboring non-hole pixels. On the other hand, if multiple MVs pass through the same pixel in F_t, we use the average of these MVs. Consequently, each pixel in F_t is associated with an MV candidate, which is referred to as the motion hypothesis with parameter k.

2. Motion Refinement: Given each parameter k Є K, we perform the block based ME to find the motion hypothesis for each pixel in find the motion hypothesis for each pixel in the intermediate frame F_t. For each parameter k, we refine the motion hypothesis field using video segmentation results. We partition the previous frame F_{t-1} and the following frame F_{t+1} jointly by employing the video segmentation technique [10]. Let S_{t-1} and S_{t+1} denote the segmentation maps of F_{t-1} and F_{t+1}, respectively, which contain the segment indices of pixels. Since each object has unique color and shape, its segment indices tend to be invariant over successive time instances. In other words, if a motion hypothesis is reliable, the segment indices of the matching pixels should be the same. Based on this observation, we introduce the unreliability detector $C_t(p, v)$ to examine the unreliability of the motion hypothesis v of pixel p in F_t, $C_t(p, v) = \; S_{t-1}(p - v) \;\; S_{t+1}(p + v)$.In other words, $C_t(p, v) = 0$ if the matching pixels $p - v$ in F_{t-1} and $p + v$ in F_{t+1} have the same segmentation index, and $C_t(p, v) = 1$ otherwise. When the motion hypothesis v of pixel p is unreliable, i.e. $C_t(p, v) = 1$, we replace it with one of neighboring reliable motion hypotheses. Note that the initial MVs obtained are used to define the temporal neighbors in the graph-based segmentation, and then the segmentation results are used to refine the MVs. Since the initial MVs are not reliable, the segmentation results may contain undesirable artifacts. To overcome this problem, we may use the refined MVs in the segmentation again and perform the refinement and the segmentation iteratively until the convergence. This iterative approach, however, would demand too high complexity and is not employed in this work.

3. Parameter Optimization: A good ME parameter should satisfy at least two properties to find an accurate MV: First, the block size should be determined so that a block contains only a single object or a group of objects with the same movement. Second, the ME direction should be chosen to avoid the deformation of objects as much as

possible. The smallest block size 8*8 preserves details in the rolling ball well, but it fails to reconstruct the texts on the calendar. On the contrary, the biggest block size 32*32 renders the texts faithfully, but it blurs the ball. In the Combination result, a higher quality frame is obtained by choosing different block sizes adaptively. Finally, we employ the graph cut technique [11] to minimize the energy function. Note that we minimize the labeling energy for ME parameters, whereas the conventional optical flow algorithms [12] minimize the energy for MVs directly.

3.5 Frame Interpolation based on Texture Optimization

The MHME algorithm estimates the MV for each pixel in an intermediate frame by choosing the ME parameter adaptively. Then use the MV field to interpolate the intermediate frame. For the interpolation, a texture optimization technique is used. It refines the MV field and the intermediate frame iteratively to correct erratic textures and achieve high interpolation quality. Furthermore, a blending technique is used, which combines multiple source pixels in the previous and following frames to reconstruct each pixel in the intermediate frame, to improve the quality further.

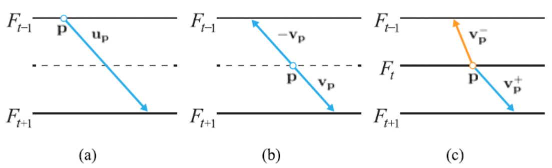


Fig 3: Comparison of the dissimilarity measures for the ME

1. Texture optimization: In the MHME in the forward or backward ME based on the matching cost J is performed which is illustrated in Fig. 3(a). Then refine the ME results to obtain symmetric MVs for the intermediate frame F_t, as shown in Fig. 3(b). However, objects sometimes exhibit complex nonlinear motions with varying velocities or are occluded by other objects. In such cases, symmetric MVs cannot describe motion trajectories faithfully. Therefore, after obtaining an initial intermediate frame, asymmetrically refine the forward MV v^+_p and the backward MV v^-_p for each pixel p. Fig. 3(c) illustrates this asymmetric motion refinement. First reconstruct an initial intermediate frame Ft using the MVs. Then, measure the texture deformation of each local patch P_p centered at p in F_t , with respect to F_{t-1} and F_{t+1}, respectively. Compute

$$J^-(v;p) = \sum_{q\in P_p} |F_t(q) - F_{t-1}(q+v)|$$
$$J^+(v;p) = \sum_{q\in P_p} |F_t(q) - F_{t+1}(q+v)| \quad (10)$$

In order to smooth the MV field, we update each V^+_p to

$$V^+_p = \arg\min\, (J^+(v_r;\bar{p})) \quad (11)$$

where $M(p)$ denotes a neighborhood of p. Perform thesetwo steps iteratively, until the MV fields converge. The texture optimization further improves the quality of the MV field by refining the MVs and the interpolated frame iteratively. Therefore, it interpolates the intermediate frame more faithfully than the previous steps.

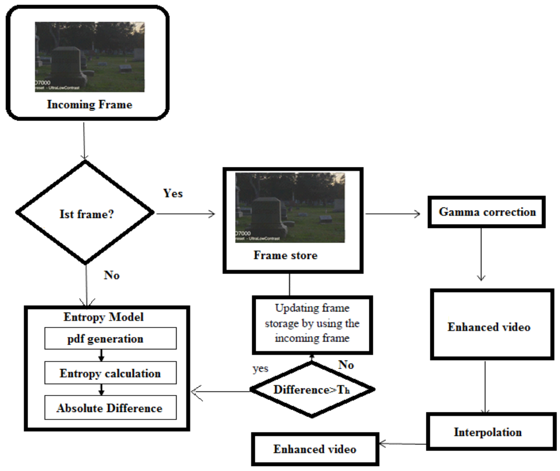
2. Blending: Let $_t(p)$ denote the synthesized pixel using the forward MV $v^+ p$ and the backward MV $v^-{}_p$. The synthesis process uses a single pixel in F_{t-1} and a single pixel in F_{t+1} to reconstruct each pixel in F_t. Finally refine $_t$ to $^*{}_t$ by blending multiple pixel values of F_{t-1} and F_{t+1} to reconstruct each pixel in $^*{}_t$.

   Using the equation:

$$\tilde{f}^*_t = (1/2)(W_{t-1} * f_{t-1} + W_{t+1} * f_{t+1}) \quad (12)$$

   Where f_t, f_{t-1}, and f_{t+1} be the vector notations for F_t, F_{t-1}, and F_{t+1}, respectively and W_{t-1} and W_{t+1} denote the blending matrices for f_{t-1} and f_{t+1}, respectively. Thus we can blend high quality intermediate frames, which preserve edge information without blurring artifacts.

Figure 4 shows the overall steps of the proposed system. The figure depicts the steps which each frame passes through to produce continuous high quality enhanced video.



4. CONCLUSION

In this paper, we present a novel framework for video enhancement. The proposed framework employs video contrast enhancement based on gamma correction with weighting distribution and subsequent motion-compensated frame interpolation based on multihypothesis motion estimation for avoiding the discontinuity of resulting video. Contrast enhancement based on gamma correction with weighted distribution composed of three major steps. First, the histogram analysis provides the spatial information of a single image based on probability and statistical inference. In the second step, the weighting distribution is used to smooth the fluctuant phenomenon and thus avoid generation of unfavourable artifacts. In the third and final step, gamma correction can automatically enhance the image contrast through use of a smoothing curve. Based on the difference of the information content, the entropy model was used to determine whether or not the transformation curve should be updated. A major problem faced by this approach is the discontinuity in the resulting enhanced video due to some missing frames. To cope up with this problem we interpolate the resulting video. Here we use a motion-compensated frame interpolation (MCFI) algorithm to increase video temporal resolutions based on multihypothesis motion estimation and texture optimization. This algorithm makes multiple motion hypotheses for each pixel using different ME parameters. Then, it solves the labeling problem to optimize the parameters and determine the best motion hypothesis for each pixel. Finally, it refines the motion hypothesis field based on the texture optimization technique and blends multiple source pixels to reconstruct each pixel in the intermediate frame. Thus the proposed method improves the brightness of dim videos and preserves the continuity of the enhanced video.

5. REFERENCES

[1] S. Huang, F. Cheng, and Yi-Sheng Chiu, ”Efficient Contrast Enhancement Using Adaptive Gamma Correction With Weighting Distribution” IEEE Trans. Image Process., vol. 22, No. 3, pp.1032–1041 Mar. 2013.

[2] Seong-Gyun Jeong,Chul Lee,and Chang-Su im,”Motion-Compensated Frame Interpolation Based on Multihypothesis Motion Estimation and Texture Optimization” IEEE Trans.Image Process., vol. 22, NO. 11, pp.4497–4509 Nov. 2013.

[3] T. Arici, S. Dikbas, and Y. Altunbasak,”A histogram modification framework and its application for image contrast enhancement,” IEEE Trans. Image Process., vol. 18, no. 9, pp.1921–1935 Sep. 2009.

[4] M. Kim and M. G. Chung, ”Recursively separated and weighted histogram equalization for brightness preservation and contrast enhancement,” IEEE Trans. Consum. Electron., vol. 54, no. 3, pp.1389–1397 Aug. 2008.

[5] Y. Kim, ”Contrast enhancement using brightness preserving bi-histogram equalization,” IEEE Trans. Consum. Electron., vol. 43, no. 1, pp.1–8 Feb. 1997.

[6] Chulwoo Lee , Chul Lee , Young-Yoon Lee , C. Kim ,”Power Constrained Contrast Enhancement for Emissive Displays Based on Histogram Equalization method” IEEE Trans. Image Process., vol. 21, pp. 80–93 Jan. 2012.

[7] B.D. Choi, J.W. Han, C.S. Kim, and S.J. Ko, ”Motioncompensated frame interpolation using bilateral motion estimation and adaptive overlapped block motion compensation,” IEEE Trans. Circuits Syst. Video Technol., vol. 17, no. 4, pp. 407–416 Apr. 2007.

[8] G. Dane and T. Q. Nguyen, ”Optimal temporal interpolation filter for motion-compensated frame rate up conversion,” IEEE Trans. Image Process., vol. 15, no. 4, pp. 978–991, Apr. 2006.

[9] V. Kwatra, I. Essa, A. Bobick, and N. Kwatra, ”Texture optimization for example-based synthesis,” ACM Trans. Graph., vol. 24, no. 3,pp. 795–802, Jul. 2005.

[10] M. Grundmann, V. Kwatra, M. Han, and I. Essa, ”Efficient hierarchical graph-based video segmentation,” in Proc. IEEE CVPR, Jun. 2010,pp. 2141–2148.

[11] Y. Boykov, O. Veksler, and R. Zabin, ”Fast approximate energy minimization via graph cuts,” IEEE Trans. attern Anal. Mach. Intell.,vol. 23, no. 11, pp. 1222–1239, Nov. 2001.

[12] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert,”High accuracy optical flow estimation based on a theory of warping,” in Proc. ECCV,May 2004, pp. 25–36.