Performance Analysis of Optical Flow: High Definition Video .vs. Regular Video

Jennifer Allen
Department of Computer Science and Information Systems
Youngstown State University
Youngstown, Ohio 44555
(937) 407-1206
jcallen@student.ysu.edu

ABSTRACT

This paper presents a preliminary result on the performance evaluation of a robust optical flow algorithm using a high definition video and a regular video. The hypothesis is that, because of its higher resolution, a high definition video can yield better optical flow data, which is critical for the motion-based face recognition research. The experiment was designed to capture the facial motion of human subjects in a video sequence, which was then broken into individual frames. A series of frame pairs of different frame intervals were then processed by an optical flow algorithm to generate dense optical flow images. Analysis of those images based on both qualitative visual examination and quantitative violation percentages indicates that a high definition video is superior to a regular video.

1. INTRODUCTION

Since the 9-11 attack on The World Trade Center in New York, homeland security and public safety have become the top priorities in the agenda of government agencies, federally funded research as well as media coverage that changed the everyday life of American people. There is an urgent need of more effective means to detect and prevent future terrorist threats. Face recognition based on computer vision and imaging technology is one of the most widely accepted and intensively studied biometrics [6, 10]. Face biometric can be used to identify a suspect at various locations, especially at the security checking points such as the entries to an airport or a public building.

Most face biometrics are developed and tested using a set of standard images in which a "clean" face is present. However, the quality of a face image acquired in a natural environment may be affected by multiple light sources. More seriously, a person may wear make-up or camouflage that drastically changes his or her appearance. All of those adverse factors can cause marked performance degradation of the conventional face recognition methods [8]. To cope with those challenges, various new face biometrics have been proposed, among which motion-based approaches that utilize either facial expressions or material properties hold great promise [3, 9, 11, 12].

Facial motion can be extracted from videos using an optical flow algorithm. The quality of an optical flow image is dependent upon the robustness of the algorithm, the complexity of the motion and the background scene, as well as the resolution of images, which in turn may affect the face recognition rate. Therefore, one way to improve the quality of optical flow results is to increase the image resolution.

The main objective of this study is to investigate whether a high definition (HD) video will produce better optic flow outputs than a regular video. The resolution of a HD video is about three times higher than that of a regular video. To ensure an objective comparison, three measures were taken: (1) all videos were acquired under the same setting; (2) since the optical flow is directly related to the frame interval, the experimental results were organized and analyzed based on a sequence of increasing frame intervals; (3) two normalized violation percentages were used to quantify the optical flow results.

To the best knowledge of the author, no previous work has been reported on comparing a HD video against a regular video in the context of optical flow performance evaluation. Although it is expected that a HD video of a higher resolution would produce a better result, no experimental data is available to support such a hypothesis. This study will provide valuable information upon which a more thorough investigation can be conducted.

The organization of this paper is as follows: Section 2 gives a brief review on the optical flow theory and its associated basic assumptions. Experimental design for video acquisition, frame extraction and flow computation is discussed in Section 3. In Section 4, the performance of a robust optical flow algorithm is then analyzed both qualitatively and quantitatively using the results obtained from HD and regular videos. Finally, conclusions are given in Section 4.

2. THEORETICAL BACKGROUND

2.1 Optical Flow

Optical flow is the velocity field that represents the three-dimensional motion of objects points across a two-dimensional image plane [7]. The fundamental governing equation to compute an optical flow field is derived from two basic assumptions: (a) the observed image brightness of an object point remains constant over a time interval of two frames; (b) the points in a small image region move in a similar speed and direction (the spatial smoothness constraint) [5].

Considering the image brightness, $E=E(x,y,t)$, as a function of image coordinates (x,y) and time (t), the brightness constancy between two frames can be expressed as a total derivative based on the conservation principle [5]:

$$\frac{dE}{dt} = 0.$$

Assuming that the brightness function is differentiable in both spatial and temporal domains, the above equation can be expanded by applying the chain rule of differentiation:

$$\frac{dE(x(t), y(t), t)}{dt} = \frac{\partial E}{\partial x}\frac{dx}{dt} + \frac{\partial E}{\partial y}\frac{dy}{dt} + \frac{dE}{dt} = 0.$$

Using the notation that $U = [u = dx/dt,\ v = dy/dt]^T$ is the motion vector, and ∇E and E_t denote the spatial and temporal gradients of the brightness function, respectively, the brightness constancy equation can be written as:

$$(\nabla E)^T U + E_t = 0.$$

The brightness constancy equation does not provide a full solution of the motion field due to various issues, such as those encountered in the well-known aperture problem. It must be further regularized by a spatial smoothness constraint:

$$obj(u,v) = (E_x u + E_y v + E_t)^2 + \lambda(u_x^2 + u_y^2 + v_x^2 + v_y^2),$$

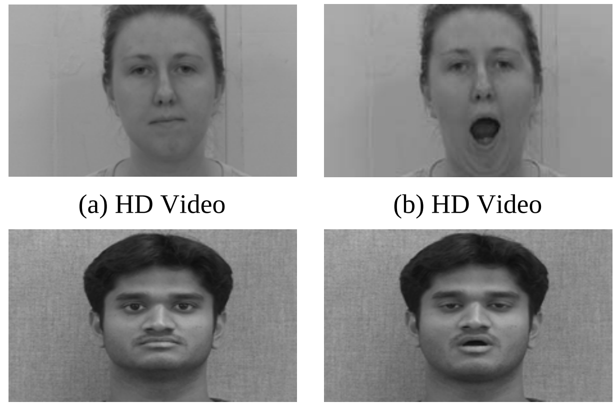
where E_x, E_y, u_x, u_y, v_x and v_y denote the partial derivatives of the corresponding variables in the x and y coordinates, and λ is the Lagrange multiplier (regularization coefficient). An optical flow solution of (u,v) obtained by minimizing the objective function represents a compromise between the observed motion and the spatial smoothness constraint.

2.2 Robust Algorithm

In real applications, the two fundamental assumptions underlying the optical flow equation are often violated, especially at the presence of multiple motions that involve physical boundaries and depth discontinuities. Various algorithms have been developed to provide a more reliable optical flow solution. Detailed discussions of those algorithms and their performance can be found in [1].

In this study, a robust algorithm based on a piecewise smoothing and multi-resolution strategy is adopted [2]. The violations of both brightness constancy and spatial smoothness are addressed in a framework that is similar to the "line-process" model for image reconstruction, where the over-smoothing of discontinuities is handled by an adaptive local method [4]. This robust algorithm has the advantage that both types of violations can be dealt with in a unified manner, which allows us to conduct a quantitative evaluation of its performance. It also handles motion boundaries well, which is beneficial to facial motion analysis.

This algorithm was implemented with the C language and is easy to use [2]. Various controlling parameters were also provided to suit different computation needs. In all the experiments, only the default parameter values were used. For example, the stage level was set to 5 and the maximum iteration number was set to 25. The algorithm converged within a reasonable time, usually less than 10 minutes on a SUN Workstation.



(a) HD Video (b) HD Video

(c) Regular Video (d) Regular Video

Figure 1. Sample frames from the HD and regular videos.

3. EXPERIMENTS

3.1 Video Acquisition

Both HD and regular videos were acquired in a normal indoor environment. Subjects sit in a chair that is about 3 meters away from the camcorder. The background was a white board. Two head-lights were placed above the head of subjects to create the desired lighting effect. The HD videos were acquired using a JVC HD Camcorder (GYHD100) with an image resolution of 1280 by 720 pixels, and the regular videos were acquired using a Canon Optura-20 Camcorder with an image resolution of 720 by 480 pixels. Both types of videos were collected at a speed of 30 frames per second. Each video sequence lasts 1-5 minutes in which a subject slowly opened and closed his or her mouth a few times. Only frontal face videos were collected.

A total of eight subjects (both male and female) participated in the experiments. To compute optical flow, individual frames were first extracted from a video sequence using the Adobe Premiere Professional tool and then were converted to the grayscale PGM format. A few sample frames from a HD video and a regular video are shown in Figure 1.

3.2 Computational Procedure

The computational procedure is designed as follows: (1) a pair of image frames was selected and then inputted into the optical flow program to produce the vertical and horizontal flow components and the violation data for the spatial smoothness constraint and the brightness constancy assumption; (2) all tests started with a pair of adjacent frames, and then with the frame interval being gradually increased until the frame interval reached nine. Based on the observation of several trial tests, the optical flow algorithm failed to yield meaningful results when the frame interval was greater than nine. The first pair was chosen when a subject started to open his or her mouth. For example, a test run would include the following pairs: (frame 10, frame 11), (frame 10, frame 12), (frame 10, frame 13), (frame 10, frame 14), (frame 10, frame 15), (frame 10, frame 16), (frame 10, frame 17), (frame 10, frame 18) and (frame 10, frame 19). Each run used the same values for all optical flow parameters.

4. RESULTS AND DISCUSSIONS

The optical flow program generates four types of image outputs: (1) vertical motion component, (2) horizontal motion component, (3) spatial smoothness violation, and (4) brightness constancy violation. The vertical and horizontal motions are displayed in the scaled 8 bits intensity images, with lower intensity (darker pixels) for smaller motion and higher intensity (brighter pixels) for larger motion. A smoothness violation image reveals where the spatial coherence constraint is violated. Smoothness violations usually occur along the motion boundaries and are shown as white pixels. Similarly, a brightness constancy image indicates where the brightness conservation assumption is violated. These optical flow images of HD videos and regular videos will be analyzed using the following rule: the less violation, the better result.

4.1 Vertical Motion

Both HD and regular videos showed good vertical motion results (see Figure 2). Large motions around the mouth (a subject was asked to open and close his or her mouth) were clearly captured as indicated by the brighter pixels, while the upper portions of the face were much darker due to less or no motion.

However, the motion field inside a face of HD videos is much smoother and more continuous than that of regular videos. For example, as shown in Figure 2 (b), the "blocking effect" is more common in regular videos than in HD videos. By "blocking effect" we mean that a block of pixels in a motion image has the same or similar intensity values. The quality of vertical motion images of HD videos is also more consistent than that of regular videos. One possible explanation is that a HD video can pick up more detailed facial movements than a regular video does.

4.2 Horizontal Motion

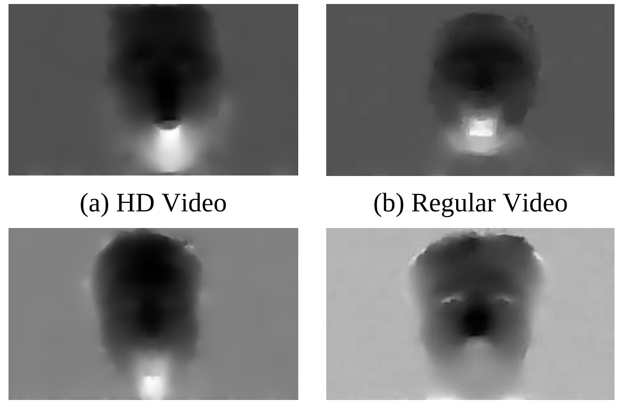
The horizontal motion images from both HD and regular videos provide less useful information due to the fact that a subject's face moved little horizontally during video acquisition (see Figure 3). The images are also relatively fuzzy and blurry. The outline of the subject's head in Figure 3 (b) is probably caused by a slight head shift between two frames.

4.3 Spatial Smoothness Violation

As expected, the violation of smoothness constraint occurred mainly around the boundaries of a moving object, such as the head, the mouth and the blinked eye (Figure 4). Almost all HD and regular videos exhibited the same violation pattern. However, one distinction is that regular videos have many violations in the background while HD videos have a relatively clean background. It is not clear what caused the background violations in regular videos, but it is likely related to the low resolution which leads to discontinuities in optical flow solutions.

4.4 Brightness Constancy Violation

The occurrence of the brightness constancy violation is much less frequent than the spatial smoothness violation, especially in HD videos. The violations usually concentrated around the mouth and the eyes as shown in Figure 5 (a). In contrast, there were more brightness violations in regular videos and they progressively got worse as the frame interval increased. Violations can also be observed in the background of regular videos (Figure 5 (b)).



(a) HD Video (b) Regular Video

(c) DH Video (d) Regular Video

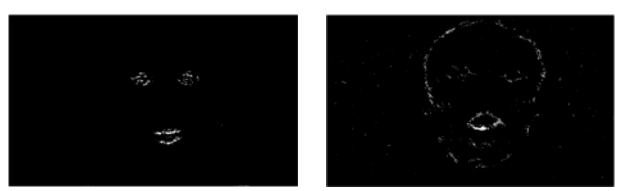
Figure 2. Examples of vertical motion component computed using the HD videos and the regular videos.

(a) HD Video (b) Regular Video

Figure 3. Examples of horizontal motion component computed using the HD videos and the regular videos.

(a) HD Video (b) Regular Video

Figure 4. Sample images of spatial smoothness violations. Note the violations in the background of regular video.



(a) HD Video (b) Regular Video

Figure 5. Sample images of brightness constancy violations.

4.5 Quantitative Comparison

The visual examination of optical flow images in the previous subsections suggests that a HD video is a better choice than a regular video. To give a more quantitative measure of the impact of image resolution on motion results, two violation percentages were computed by taking the ratio of the number of pixels with violation to the number of pixels without violation, i.e. the ratio of the number of white pixels to the number of black pixels in Figure 4 and Figure 5. The two violation percentages allow for more objective comparison of HD videos against regular videos on a normalized basis.

Other than the image resolution, the frame interval between a frame pair also has a significant impact on the quality of optical flow results. For example, as shown in Figure 6, a pair of nine frame interval has much severe smoothness violations than a pair of one frame interval. The violation started with the eyes and gradually propagated to the regions around the mouth as the subject opened his/her mouth to the maximum degree.

To assess the two factors simultaneously (image resolution and frame interval), a series of violation percentages were computed using image pairs of increasing frame intervals (from 1 to 9). The average results of all video sequences are plotted in Figure 7 and Figure 8. Three important observations can be made: (1) HD videos constantly show better performance than regular videos as measured by the low violation percentages, regardless of the value of frame interval; (2) a quasi-monotonic relationship exists between the violation percentages and the frame interval. In other words, the quality of optical flow results deteriorates as the frame interval increases. This is true for both HD videos and regular videos; (3) brightness constancy violation is less affected by the frame interval, especially in HD videos as indicated by the relatively flat curve in Figure 8. This is partially due to the simple mouth motion and the indoor lighting.

(a) One Frame Interval (b) Nine Frame Interval

Figure 6. Spatial smoothness violations with different frame intervals in a HD video.

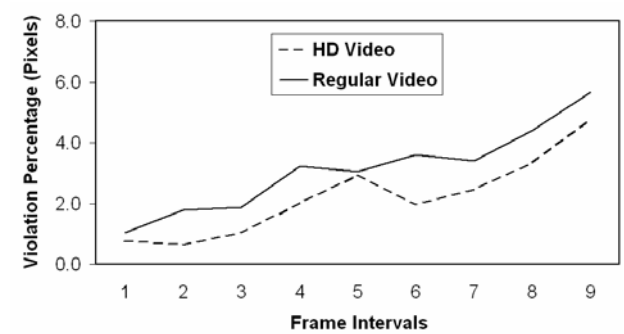


Figure 7. The relationship between the spatial smoothness violation percentage and the frame interval.

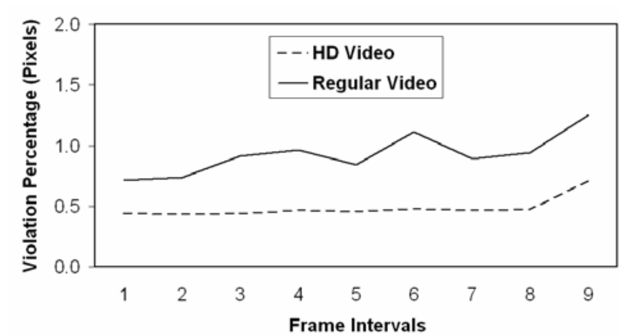


Figure 8. The relationship between the brightness constancy violation and the frame interval.

5. CONCLUSIONS

High quality motion data extracted from digital videos is essential for a wide variety of applications, such as face recognition, multimedia database, surveillance and security, robotic navigation and realistic animation. This study investigated the feasibility of using high definition videos to provide better optical flow results. The experiment was carried out by comparing the outputs of a robust optical flow algorithm using HD videos and regular videos. Both the qualitative visual examination (vertical and horizontal motion components) and the quantitative analysis (the percentage of spatial smoothness violation and the percentage of brightness constancy violation) suggest that HD videos outperformed regular videos.

It should be stressed that the above conclusions are drawn from a preliminary experiment with a limited set of video samples in a well controlled environment (slow mouth motion with a default capture speed of 30 frames per second). Further investigations with a larger data set in a more sophisticated setting such as the outdoor lighting condition are needed in order to make a more statistically sound statement.

6. ACKNOWLEDGMENTS

This work has been supported in part by the National Science Foundation under grant No. 0453463 (Research Experience for Undergraduates) at University of South Florida, and the URC grant 06-07 #16-07 of Youngstown State University. The author is deeply grateful to Dr. Dmitry Goldgof for his mentoring. The author also thanks Dr. Miguel Labrador for organizing such a wonderful undergraduate research program. The author is also indebted to Mr. Vasant Manohar, Dr. Yong Zhang and CSIS Chair Thomas Bodnovich at Youngstown State University for their continuous encouragement and support.

7. REFERENCES

[1] Barron, J. L., Fleet, D. J., Beauchemin, S. S., and Burkitt, T. A., Performance of optical flow techniques, *Proceedings of Computer Vision and Pattern Recog.*, pp. 236-242, 1992.

[2] Black, M. J., and Anandan, P., The robust estimation of multiple motions: parametric and piecewise-smooth flow fields, *Computer Vision and Image Understanding*, 63(1), pp. 75-104, 1996.

[3] Essa, I. and Pentland, A., Coding, analysis, interpretation and recognition of facial expressions, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7), pp. 757-763, 1997.

[4] Geman, S., and Geman, D., Stochastic relaxation, Gibbs distributions and Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(5), pp. 401-412, 1991.

[5] Horn, B. K. P., and Schunck, B. G., Determining optical flow, *AI Memo 572*, Massachusetts Institute of Tech., 1980.

[6] Wayman, J., Jain, A. K., Maltoni, D., and Maio, D., *Biometric Systems: Technology, Design and Performance Evaluation*, Springer Verlag, 2005.

[7] Kearney, J. K., and Thompson, W. B., Bounding constraint propagation for optical flow estimation, in *Motion Understanding* (Aggarwal, J. K., and Martin, W., editors), Kluwer, Boston, 1988.

[8] Phillips, P. J., Flynn, P. J., Scruggs, W. T., Bowyer, K. W., Chang, J., Hoffman, K., Marques, J., Min, J., and Worek, W. J., Overview of the face recognition grand challenge, *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 947-954, 2005.

[9] Shah, M., and Jain, R., *Motion-Based Recognition*, Academic Publishers, Netherlands, 1997.

[10] Turk, M., and Pentland, A., Eigenfaces for recognition, *Journal of Cognitive Neuroscience*, (3)1, pp. 71-86, 1991.

[11] Manohar, V., *Video-Based Person Identification Using Facial Strain Maps as a Biometric*, M.S. Thesis, Dept. of Computer Science and Engineering, University of South Florida, 2006.

[12] Zhang, Y., Kundu, S. J., Goldgof. B. D., Sarkar, S., and Tsap, L. V., Elastic face, an anatomy-based biometrics beyond visible cue, *Proceedings of the 17th International Conference on Pattern Recognition (ICPR-2004)*, vol. 2, pp. 19-22, 2004.